UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUÍMICAS

TESIS DOCTORAL

Resonancia magnética nuclear aplicada al estudio sobre el reconocimiento molecular de ligando con fucosa mediado por el receptor DC-SIGN

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Pablo Valverde Sánchez

Directores

Jesús Jiménez Barbero
Francisco Javier Cañada Vicina
Niels-Christian Reichardt

Madrid

© Pablo Valverde Sánchez, 2020

FACULTAD DE CIENCIAS QUÍMICAS

TESIS DOCTORAL

RESONANCIA MAGNÉTICA NUCLEAR APLICADA AL ESTUDIO SOBRE EL RECONOCIMIENTO MOLECULAR DE LIGANDOS CON FUCOSA MEDIADO POR EL RECEPTOR DC-SIGN

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

PABLO VALVERDE SÁNCHEZ

DIRECTOR

JESÚS JIMÉNEZ BARBERO
JAVIER CAÑADA VICINAY
NIELS-CHRISTIAN REICHARDT

© PABLO VALVERDE SÁNCHEZ

FACULTAD DE CIENCIAS QUÍMICAS

RESONANCIA MAGNÉTICA NUCLEAR APLICADA AL ESTUDIO SOBRE EL RECONOCIMIENTO MOLECULAR DE LIGANDOS CON FUCOSA MEDIADO POR EL RECEPTOR DC-SIGN

***

PABLO VALVERDE SÁNCHEZ

TESIS DOCTORAL
2020

Dirigida por:
JESÚS JIMÉNEZ BARBERO
FRANCISCO JAVIER CAÑADA VICINAY
NIELS-CHRISTIAN REICHARDT

AGRADECIMIENTOS

Quiero agradecer a los organismos que ha hecho posible la realización de este trabajo, en especial al Ministerio de Educación, Cultura y Deporte; así como a los centros en los que se ha llevado a cabo el desarrollo de experimentos, CIC bioGUNE (Bizkaia), CIB CSIC (Madrid) y la Universidad de Southampton (UK).

A mi familia, amigos y compañeros de trabajo.

CONTENTS

ABBREVIATIONS AND ACRONYMS

APC	Antigen Presenting Cell
ASGR	ASialoGlycoprotein Receptor
BgA	Blood group A antigen
BgB	Blood group B antigen
BMRB	Biological Magnetic Resonance Bank
BSA	Bovine Serum Albumin
CLR	C-type Lectin-like Receptors
CORCEMA-ST	COmplete Relaxation and Conformational Exchange MAtrix - Saturation Transfer
COSY	COrrelation SpectroscopY
CPMG	Carr-Purcell-Meibom-Gill
CRD	Carbohydrate Recognition Domain
CSA	Chemical Shift Anisotropy
CSP	Chemical Shift Perturbation
CTLD	C-Type Lectin-like Domain
DAMP	Damage-Associated Molecular Pattern
DAST	DiethylAminoSulfur Trifluoride
DC	Dendritic Cell
DCIR	Dendritic Cell ImmunoReceptor
DCM	DiChloroMethane
DC-SIGN	Dendritic Cell-Specific ICAM-3-Grabbing Non-integrin
DC-SIGNR	Dendritic Cell-Specific ICAM-3-Grabbing Non-integrin Related (also **L-SIGN**)
DENV	DENghe Virus
DIAD	DiIsopropyl AzoDicarboxylate
DMF	DiMethyl Formamide
DMP	Dess-Martin Periodinane
DOSY	Diffusion Ordered SpectroscopY
DTT	DiThioThreitol
EBOV	EBOla Virus
ECD	ExtraCellular Domain
EDTA	Ethylene Diamino Tetraacetic Acid
ELISA	Enzyme-Linked ImmunoSorbent Assay
EM	Electron Microscopy
FDC	Follicular Dendritic Cell
FID	Free Induction Decay
FT	Fourier Transform *or* FucosylTransferase

GnT	GlycosylTransferase
GOase	Galactose Oxidase
GT	GalactosylTransferase
hCMV	human CitoMegaloVirus
HIV	Human Immunodeficiency Virus
HMBC	Heteronuclear Multiple Bond Correlation
HMQC	Heteronuclear Multiple Quantum Correlation
HRMS	High Resolution Mass Spectrometry
HSQC	Heteronuclear Single Quantum Correlation
HTS	High Throughput Screening
IAV	Influenza A Virus
IPTG	IsoPropyl β-D-1-ThioGalactopyranoside
ITC	Isothermal Titration Calorimetry
KSHV	Kaposki´s Sarcoma associated Herpes Virus
LAM	LipoArabinoMannan
LB	Lysogeny Broth
LC	Langerhans Cell
LDN	Di-N-acetyllactose
LDN-DF	LacDiNAc DiFucosylated
LDNF	LacDiNAc Fucosylated
Le	Lewis-type/Lewis antigen
LN	N-acetyllactose
LNFP III	Lacto-N-FucoPentaose type III
LOGSY	Ligand Observed via Gradient SpectroscopY
L-SIGN	Dendritic Cell-Specific ICAM-3-Grabbing Non-integrin Related (Also **DC-SIGNR**)
MBL/MBP	Mannose Binding Lectin/Protein
MCL	Macrophage C-type Lectin
MD	Molecular Dynamics
MGL	Macrophage Galactose-type Lectin
MHC	Major Histocompatibility Complex
MMR	Macrophage Mannose Receptor
NK	Natural Killer
NMR	Nuclear Magnetic Resonance
NOD	Nucleotide-binding Oligomerization Domain
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect SpectroscopY
NUS	Non-Uniform Sampling
OD	Optical Density

PAMP	Pathogen-Associated Molecular Pattern
PCS	PseudoContact Shifts
pDC	Plasmacytoid Dendritic Cell
PDT	1,3-PropaneDiThiol
PRR	Pathogen Recognition Receptor
RCI	Random Coil Index
RIG	Retinoic acid-Inducible Gene
ROESY	Rotating-frame nuclear Overhauser Effect SpectroscopY
SAR	Structure-Activity Relationship
SCD	Sickle Cell Disease
SDS-PAGE	Sodium Dodecyl Sulfate–PolyAcrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SPR	Surface Plasmon Resonance
SRCL	Scavenger Receptor C-type Lectin
ST	SialylTransferase
STD	Saturation Transfer Difference
STDD	Saturation Transfer Difference Difference
TALOS+	Prediction of Protein Backbone & Sidechain Torsion Angles from NMR Chemical Shifts
TBAB	TetraButylAmino Bromide
TBAF	TetraButylAmino Fluoride
TCCA	TriChloroisoCyanuric Acid
TEMPO	2,2,6,6-TEtraMethylPiperidin-1-yl)Oxyl
THF	TetraHydroFuran
TLR	TOLL-Like Receptors
TOCSY	TOtal Correlation SpectroscopY
TROSY	Transverse Relaxation Optimized SpectroscopY
UV	UltraViolet spectroscopy

SUGARS

Fuc/L-Fuc	L-fucopyranose
Gal/D-Gal	D-galactopyranose
GalNAc/D-GalNAc	N-acetyl-D-galactopyranose
Glc/D-Glc	D-glucopyranose
GlcNAc/D-GlcNAc	N-acetyl-D-glucopyranose
Man/D-Man	D-mannopyranose
ManNAc	N-acetyl-D-mannopyranose
OMe-D-Gal	Methyl D-galactopyranoside
OMe-D-GalNAc	Methyl N-acetyl-D-galactopyranoside
OMe-L-Fuc	Methyl L-fucopyranoside

AMINO ACIDS

Ala/A	Alanine
Arg/R	Arginine
Asn/N	Asparagine
Asp/D	Aspartic acid
Cys/C	Cysteine
Gln/Q	Glutamine
Glu/E	Glutamic acid
Gly/G	Glycine
His/H	Histidine
Ile/I	Isoleucine
Leu/L	Leucine
Lys/K	Lysine
Met/M	Methionine
Phe/F	Phenylalanine
Pro/P	Proline
Ser/S	Serine
Thr/T	Threonine
Trp/W	Tryptophan
Tyr/Y	Tyrosine
Val/V	Valine

Abstract

NMR INSIGHTS INTO DC-SIGN RECOGNITION OF FUCOSYLATED ANTIGENS

This dissertation is focused on the study of the human protein DC-SIGN, a key lectin involved in the development of diverse pathologies, paying attention to the conformational, dynamics and energy aspects that modulate its interaction with carbohydrate ligands. In particular, the present work has exhaustively analyzed recognition events involving ligands that contain the natural monosaccharide **L-fucose**.

The interest in studying this protein has extensively grown since the beginning of this century. DC-SIGN is a membrane receptor within the C-type lectin family that is directly involved in diverse infectious processes, being HIV one of the most relevant examples described at first. Moreover, it has been also demonstrated its implication **in several other pathologic processes involving many different organisms**, namely viruses (Ebola), bacteria (*M. tuberculosis*), fungi (*C. albicans*) and parasites (*S. mansoni*). DC-SIGN recognizes those oligosaccharides that coat the external membrane or envelope of the cited pathogens. The range of oligosaccharides that can be recognized by this lectin is fairly wide, although most of them always contain the monosaccharides **D-mannose (Man) and L-fucose (Fuc)**, which directly interact at the primary binding site. The recognition of D-mannose had been widely investigated at a molecular level, especially by using X-Ray crystallography. Thus, this work has concerned the interaction with Fuc-containing antigens, essentially at describing the **recognition of the histo blood group antigens A and B** from a molecular point of view. These ligands display similar scaffolds with two terminal sugars, one common (L-fucose), and other that is different in each case: D-galactose (Gal, B antigen) or N-acetyl-D-galactosamine (GalNAc, A antigen). Considering the complexity of the system, the project has been undertaken by using Nuclear Magnetic Resonance (NMR) experiments, which have provided essential data both from the receptor´s perspective and from the ligand´s one.

From the gathered information, a **suitable and detailed model of interaction between DC-SIGN and both antigens** has been proposed, which enables explaining their slightly enhanced affinities in comparison with that of a single Fuc moiety. Moreover, the results reasonably fit with those previously reported data from solid phase arrays and have highlighted relevant aspects on the recognition these antigens when compared to the Lewis-type related ones. In particular, it has been reported that the stability of the LewisX/DC-

SIGN complex is contributed by the interaction of the terminal D-galactose at a secondary site, as well as a stable hydrophobic contact formed between the N-acetyl group of the central sugar (N-acetyl-D-glucosamine) and a close protein residue (Val351). It has been found herein that the terminal D-galactose in the B antigen actually plays the opposite role. Namely, it establishes favorable aliphatic contacts with the nearby Val351, whereas no significant interactions are observed at the secondary site. These features also explain why Lewis-type antigens are usually slightly better binders of DC-SIGN that the blood group determinants.

Finally, the crucial role played by the protein scaffold and the extended structure of the studied glycan have been scrutinized as well. Langerin, a homologue receptor of DC-SIGN, shows a drastically different interaction mode with the blood group antigen B than that demonstrated herein for DC-SIGN. Indeed, the terminal Gal residue points towards the secondary site rather than establishing aliphatic contacts as in DC-SIGN. Additional **differences have also been found out between the A and B antigens** in terms of dynamics. Specifically, our NMR studies have revealed that the B antigen, but not the A, can be targeted by DC-SIGN through its terminal Gal moiety. On one hand, this finding evidences the low specificity already known for this type of receptors and proves the underlying complexity in developing selective inhibitors for this particular lectin. On the other hand, it points out how subtle structural differences are at the heart of the found different affinities and dynamic behavior of the ligands, hence encouraging the search for a global perspective to jointly understand all these factors and efficiently apply them for drug design.

As already mentioned, DC-SIGN selectivity is rather broad, a common feature shared by the lectins within the same family. However, this wide selectivity usually overlaps with that of other lectins, but does not exactly match. For DC-SIGN, understanding such a selectivity have aided at determining the mode in which this lectin is able to recognize the B antigen through the terminal Gal. Specifically, ^{19}F-NMR data have allowed to evaluating the required geometric features for the efficient recognition of a given monosaccharide. Concerning L-Fuc, three fluorinated analogues (at C2, C3 and C4, respectively), have been evaluated, assessing that this monosaccharide is exclusively recognized through its hydroxyls groups at C3 and C4, but not through OH2 and OH3 as reported for other related lectins (e.g. langerin and MBP). Moreover, one of the used fluorofucoses (**3F-Fuc) is not commercial and has been *de novo* synthesized** as a part of the thesis project.

Finally, additional recognition studies have been performed for other **two representative Lewis-type oligosaccharides** found in the glycoproteins from the parasite *S. Mansoni*. In particular, the LDNF antigen presents an structure very similar to that of the "Lewis X" antigen, except for the presence of GalNAc at its end rather than Gal. Instead, the other antigen, the LDN-DF, essentially contains the LDNF trisaccharide along with an additional L-Fuc residue.

Our NMR-based methodology, in combination with computational tools, has provided key data to propose the 3D recognition models for both molecules. Interestingly, the recognition of LDNF takes place through a binding pose that is remarkably similar to that described for the "Lewis X" antigen. In contrast, in spite of including two L-Fuc residues in its structure, LDN-DF displays a very poor affinity for DC-SIGN, which is related to the particular 3D arrangement of the sugars forming its structure: the interaction with L-Fuc places the surrounding sugar residues away from the binding site so that their positive contribution to the binding disappear. Once more, these results highlight how the available 3D-models provide essential information for explaining and predicting the recognition features of a given oligosaccharide by a biological receptor.

Resumen

RESONANCIA MAGNÉTICA NUCLEAR APLICADA AL ESTUDIO SOBRE EL RECONOCIMIENTO MOLECULAR DE LIGANDOS CON FUCOSA MEDIADO POR EL RECEPTOR DC-SIGN

La presente tesis doctoral se dirige al estudio de la proteína humana DC-SIGN desde una perspectiva estructural, atendiendo a aquellos aspectos que, a nivel molecular, determinan su interacción con sus ligandos de tipo carbohidrato (conformación, interacciones intermoleculares (vdW)…). En concreto, el trabajo aquí expuesto se ha enfocado en el análisis de los procesos de reconocimiento en los que participan ligandos que contienen el monosacárido natural **L-fucosa** en su estructura.

El interés por estudiar dicha proteína se ha acrecentado desde comienzos de este siglo, cuando DC-SIGN fue definitivamente identificada como un receptor de membrana de la familia de las *C-type lectins* y, más importante, se demostró su implicación directa en los procesos de infección por VIH. A partir de aquí, sucesivos trabajos permitieron subrayar la **importancia de este receptor en otros procesos infecciosos que incluían patógenos de todo tipo**, ya fueran virus (Ébola), bacterias (*M. tuberculosis*), hongos (*C. albicans*) o parásitos (*S. mansoni*). Desde un punto de vista estructural, es bien sabido que DC-SIGN reconoce e interacciona con oligosacáridos presentes en la superficie externa de estos agentes patógenos. El abanico de oligosacáridos que esta proteína es capaz de reconocer es bastante amplio, aunque es bien sabido que todos ellos contienen fundamentalmente los monosacáridos **D-manosa y L-fucosa**, con los que interacciona directamente. La interacción con D-manosa ha sido ampliamente estudiada a nivel molecular, especialmente mediante cristalografía de rayos X, lo cual ha permitido disponer de varios modelos que describen con un alto grado de detalle cómo ocurre el reconocimiento de este azúcar y cómo influyen tanto la estructura de la proteína como la del propio oligosacárido en dicho reconocimiento. En cambio, aunque la interacción de DC-SIGN con L-fucosa ha sido estudiada en paralelo, mucha de la información existente referente al reconocimiento de oligosacáridos fucosilados no está completamente justificada desde un punto de vista estructural, lo que ha dificultado en muchos casos la búsqueda de patrones estructurales que permitan entender bien cómo tiene lugar el reconocimiento.

Por este motivo, la tesis se ha centrado fundamentalmente en describir desde un punto de vista molecular el **reconocimiento de los determinantes antigénicos de la sangre A y B**, cuyas estructuras son similares y contienen

dos azúcares terminales, uno común, que es la L-fucosa, y uno que los diferencia, que es la D-galactosa para el antígeno B y la N-acetil-D-galactosamina para el A. Existe una limitada cantidad de información acerca de cómo tiene lugar el reconocimiento de los antígenos de tipo *Lewis*, diferentes en cuanto a composición y organización de los enlaces entre residuos. Sin embargo, no hay ningún trabajo precedente que haya descrito el reconocimiento molecular de los antígenos que en esta tesis se han estudiado.

A través de la información recopilada por medio de experimentos de Resonancia Magnética Nuclear (RMN), se han conseguido **modelos 3D de la interacción entre DC-SIGN y ambos antígenos**, que han permitido explicar por qué estas estructuras poseen una afinidad ligeramente superior a la exhibida por la L-Fuc, y concuerdan con los datos previamente publicados mediante ensayos en fase sólida. Además, los modelos propuestos han permitido subrayar aspectos relevantes acerca del reconocimiento de antígenos de este tipo y compararlos con los antígenos de tipo *Lewis*. Así, para el antígeno "Lewis X" hay una doble contribución a la estabilidad, que viene tanto de la D-Gal terminal que interacciona en el sitio secundario como de la GlcNAc central, que forma una interacción hidrofóbica estable con un residuo cercano de la proteína (Val351) a través de su grupo acetilo. En cambio, se ha podido demostrar que el rol de estabilizar dicha interacción alifática es asumido por la D-Gal terminal en el caso de los ligandos aquí estudiados, mientras que la importancia de las interacciones en el sitio secundario es menor. Además, se han encontrado **diferencias entre los propios antígenos A y B** en cuanto al reconocimiento desde un punto de vista dinámico. Los estudios de RMN han revelado que el determinante B, pero no el A, puede unirse a DC-SIGN a través de la D-Gal terminal, lo cual pone de manifiesto por un lado la baja especificidad de este receptor y, por ende, aumenta la dificultad de desarrollar inhibidores exclusivos para esta lectina. Por otro lado, este trabajo señala lo sutiles que pueden ser las variaciones estructurales que hacen que las afinidades y el comportamiento de ligandos similares sean diferentes y lo esencial que resulta entender el conjunto de factores que gobiernan estos cambios para explotarlos en beneficio del desarrollo de fármacos.

Adicionalmente, la selectividad de DC-SIGN se ha estudiado usando técnicas de ^{19}F-RMN con las que se han descifrado los requerimientos estructurales necesarios para que un azúcar simple sea reconocido por DC-SIGN. En el caso de la L-Fuc, se emplearon tres análogos fluorados (En C2, C3, y C4, respectivamente). Sorprendentemente, se encontró que este monosacárido solo puede ser reconocido a través de sus grupos hidroxilo en C3 y C4, pero no

a través de C2 y C3 como se ha observado en otras lectinas (ej. langerina y MBP). Además, como parte de la tesis y en colaboración con la Universidad de Southampton, se ha desarrollado un **proceso de síntesis completamente nuevo para obtener la L-Fuc fluorada en C3**, la única de las tres Fuc fluoradas cuya síntesis no se había descrito.

Finalmente, el trabajo principal de la tesis se ha expandido al estudio de **dos antígenos de tipo *Lewis*** representativos que se encuentran formando parte de las glicoproteínas del parásito *S. mansoni* a lo largo de sus diferentes estadios de desarrollo (LDNF y LDN-DF). De forma análoga, las técnicas de RMN se han aplicado a la descripción de esta interacción con ayuda de técnicas computacionales. Curiosamente, el reconocimiento del LDNF tiene lugar de forma similar al del antígeno "Lewis X". En cambio, el LDN-DF, a pesar de poseer dos residuos de Fuc en su estructura, posee una afinidad notablemente baja debido a cómo estos residuos están ordenados en la estructura 3D, lo que hace que la contribución de los azúcares vecinos sea despreciable y la estabilización positiva desaparezca. Estos resultados destacan, una vez más, la importancia de disponer de modelos tridimensionales fidedignos para evaluar y predecir los factores geométricos que determinan si la estructura de un oligosacárido es adecuada para su reconocimiento por un receptor biológico.

CHAPTER 1: INTRODUCTION

1.1. The immune response

Our body is a complex system of specialized cells organized in different tissues and organs. All of them, together with nervous and circulatory systems, work in coordination to ensure the individual surveillance, establishing a correct balance that is termed **homeostasis**. However, there is an endless battle between our cells and a plethora of foreign organisms that continuously threaten such a crucial balance. Those organisms, which include viruses, bacteria, parasites and fungi, are commonly named as **pathogens**, and the group of specific cells, tissues and organs in charge of attacking and eliminating these threads form what we jointly know as the **immune system.** Apart, disruption of normal homeostasis does not have to be necessarily triggered by external entities; sometimes the alterations come also from injured or damaged tissues (trauma), or from malfunctioning cells (tumoral cells), and the immune system takes actions to control these as well [1].

The immune cells are collectively called **leucocytes**, and are divided into several subsets of cells primarily distinguished by their roles. Actions to eradicate pathogens or restore normal tissue operation can be carried out directly by leucocytes, and this clearance process is referred to as **cell-mediated immunity**. In addition, certain specialized leucocytes (B cells) produce soluble molecules (**antibodies**), which directly act as effectors without cell intervention, and this process is called **humoral immunity** [2].

Commonly, the immune responses are classified in two main types that essentially differ from each other in two basic aspects: the mechanism to detect the thread and the speed of the response [2-3]:

- The **innate immune response** is immediately triggered within minutes or hours after loss of the correct homeostatic state. The cell types included in this group are dendritic cells (DCs), macrophages, mast cells, granulocytes (neutrophils, basophils and eosinophils) and natural killer cells (NK cells). These cells are often present in large amounts and are spread out over most of peripheral tissues, thus activating quick mechanisms to attack, repair, and/or recruit other cells for managing infections. Besides, the innate immune system includes pre-existing physiological barriers (skin, gastric acids, secreted enzymes) that hamper pathogen entry, and other enzymes present in the blood which are part of the complement system (proteases, antimicrobial peptides). Once activated, these enzymes are aimed at breaking down bacterial membranes.

At the molecular level, the innate response is initiated by **recognition processes** involving the patrolling cells and certain molecular components that become accessible in the extracellular milieu, either released or exhibited by damaged/dead host cells **(damage-associated molecular patterns, DAMPs)**, or presented in the surface of the invading organisms **(pathogen-associated molecular patterns, PAMPs)**. In both cases, the recognition is mediated by **pattern recognition receptors (PRRs)**, fixed to the plasma membrane or free as soluble receptors in the extracellular environment. To note, DAMPs and PAMPs are usually common molecular signatures shared by many types of host cells and pathogens, and PRRs are therefore adapted to recognize these multiple patterns, conferring to innate immune cells the property of broad recognition. This feature also enables a faster detection of potential dangers.

- The **adaptive immune response** is driven by cells named **lymphocytes,** which are subdivided in B lymphocytes and T lymphocytes. This response takes much longer to become effective, often few days, in part because lymphocytes are present in low numbers and located at lymphoid tissues, where they need to be firstly stimulated by other leukocytes. Once activated, these cells must proliferate to eventually promote protective effects, and consequently, the overall response is delayed. However, adaptive responses are crucial because they give rise to **immunological memory**, a key aspect of the immune response that strengthen our defenses in case of a repeated exposure to the same thread and allow us to face the evolving mechanisms of harmful pathogens to evade or block the immune responses. In this regard, lymphocytes make use of processes like somatic recombination to produce almost unlimited amounts of different **antigen receptors** that are highly specific. Thus, each antigen receptor is designed to exclusively recognize one molecular component present in a specific pathogen (**antigen**), in contrast to PRRs. Those lymphocytes bearing the suitable receptors to target an invading pathogen are selected and develop into short-lived effector cells and long-lived memory cells. The first wave of effector cells initiate a primary immune response, whereas the memory cells will help the organism to trigger a secondary immune response: a faster and stronger wave of attacking cells in case of a future re-infection (Figure 1.1).

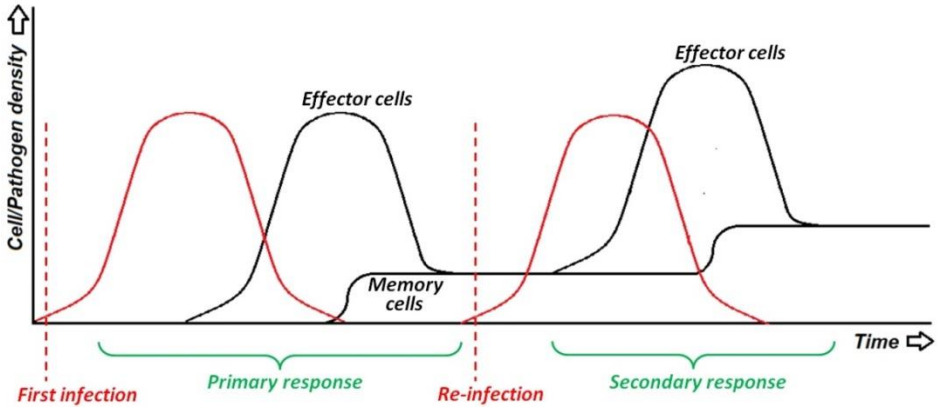


Figure 1.1. *Schematic evolution of an infection process over time. When the density of foreign microorganisms overpasses a certain threshold level, the innate response is initiated (within the first 24 hours after the pathogen entry). The adaptive response starts much later (about 72-96 hours later), and the effector cells are in charge of completely neutralizing the thread (black profile). Once finished, a percentage of memory cells are maintained, triggering a faster global immune response after a subsequent re-infection.*

Although this classification method is robust and allows to distinguishing leucocytes according to their roles, it is important to highlight that the immune cascade is not strictly sequential: every immune response always encompasses the two mechanisms working at once, the innate and the adaptive, and this interplay has a beneficial synergic effect in the host defense (Figure 1.2). When a pathogen surpasses the physical barriers and the homeostasis is endangered, complement activation can take place to contain the infection. Simultaneously, macrophages, DCs and granulocytes are continuously sampling the local microenvironments in search of PAMPs by means of their PRRs. In case of infection, some cells of the innate immune system, like granulocytes, directly fight the menace through phagocytosis. In parallel, other cells like DCs are engaged in recognizing and taking up the entire pathogen or pathogen debris to the lymph nodes, where they stimulate the response mediated by lymphocytes [4]. While some lymphocytes are developing into memory cells, others are recruited to the sites of inflammation and act as effector cells, helping to kill microbes, or alternatively activating signaling pathways to stimulate new lymphocytes and cells from the innate immune system.

In this “dialogue” between cells from innate and adaptive systems, **cytokines** play a crucial role in establishing intercellular cross-talk. They constitute a wide family of peptides that leukocytes use to send information about the

state of the tissues and the presence of external pathogens. Particularly, chemokines are a subset of relevant cytokines that immune cells employ to promote migration of other cells to the sites of actuation. Dendritic cells and macrophages, as well as CD4 T cells, are actually considered as the key mediators between the innate and the adaptive response (Figure 1.2), since these subtypes are specialized in cell-cell communication and regulatory functions in immunity [5].

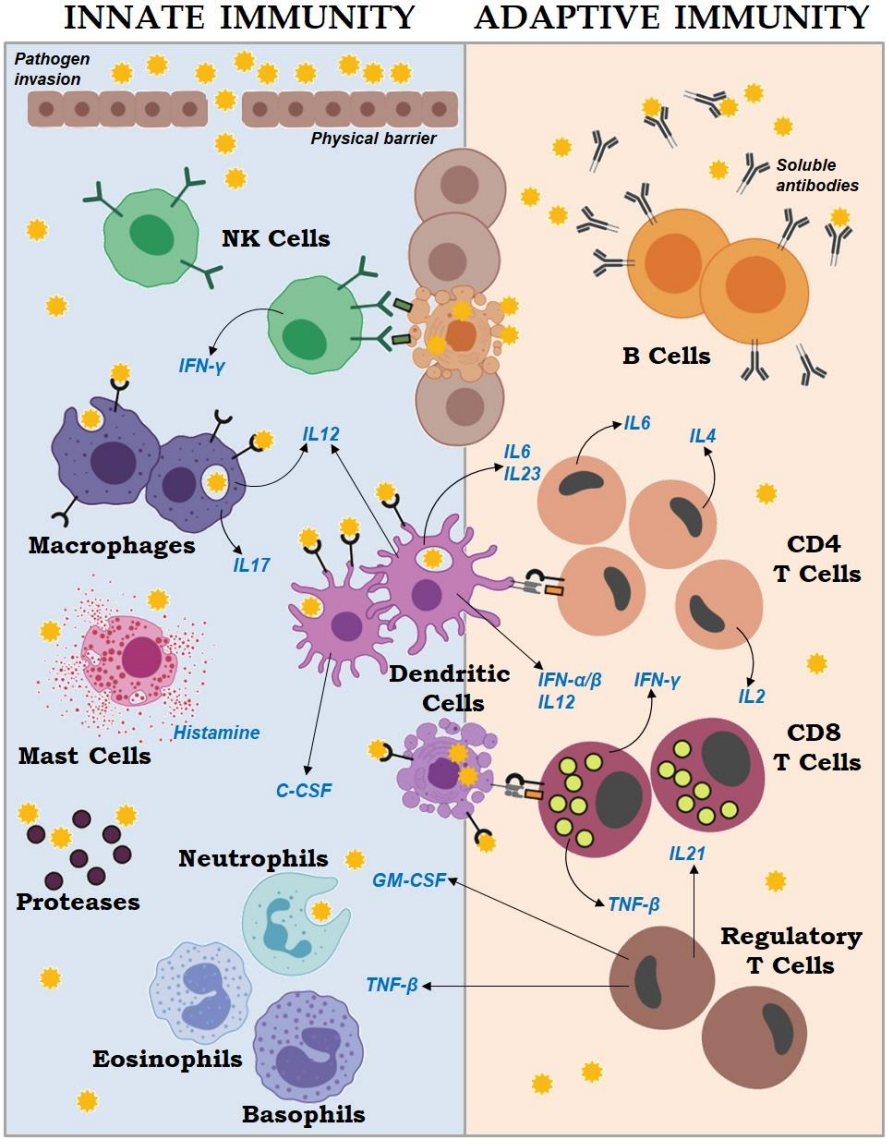


Figure 1.2. *Common leukocytes of the immune system and their responses against an ongoing infection. The arrows with the blue labels represent some different cytokines secreted as a result of the interaction with a pathogen or activation via another cytokine produced before. Adapted from [5].*

1.1.1. Cells of the immune system

All cells from the immune system share a common origin. **Stem cells** are the earliest blood cells, which will subsequently differentiate into different precursors (myeloid and lymphoid) to ultimately originate various subtypes of leucocytes, as well as other blood cells (erythrocytes, platelets). Myeloid precursors will develop into most of the blood cells and innate leucocytes (Figure 1.3). The later ones will distribute throughout the entire organism through the blood circulation and the lymphatic system. In contrast, lymphoid precursors will give rise to B and T lymphocytes at the primary lymphoid tissues (bone marrow and thymus, respectively (Figure 1.3)). **Dendritic cells (DCs)** are a particular case since some subsets derive from myeloid precursors and other from lymphoid ones [1][6-7]. Specific details about DCs will be addressed to a larger extent in the next section.

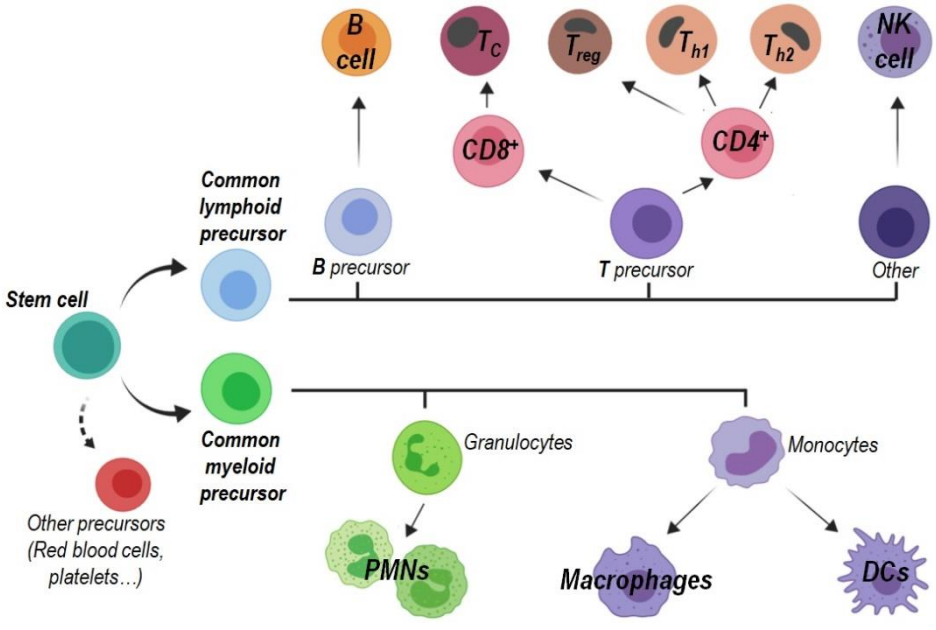


Figure 1.3. *Origin of the leukocytes. To note, most of the DCs come from myeloid precursors, although a certain DC lineage has been evidenced to have a lymphoid precursor (not represented).*

The main innate leucocytes, apart from **dendritic cells**, are macrophages, mast cells, neuthrophils, basophils, eosinophils, and natural killer cells (NK cells) [5,8] (Figure 1.2):

- **Macrophages** reside in all tissues and are relatively long-lived. They are capable of removing harmful microorganisms by phagocytosis.

Additionally, they release cytokines and chemokines to promote local inflammation and cell recruitment.

- **Neutrophils** are the most abundant leucocytes in the bloodstream, with a quite short life (up to 5 days). As macrophages, neutrophils exhibit a great capacity to eat and destroy entire pathogens by phagocytosis, using preexisting toxic components stored in specialized granules. Also, they can undertake the formation of neutrophil extracellular traps (NETs).
- **Mast cells** are fixed to mucosal and connective tissues, where they act as sentinels, detecting tissue injuries or pathogens. They are aimed at triggering inflammation through the release of cytokines and histamine, especially in response to pathological and allergic processes.
- **Eosinophils** are involved in analogous inflammatory responses to allergies, but in contrast to mast cells, they can be recruited to the affected areas. They may also perform particular immune processes to fight parasitic infections.
- **Basophils** often support eosinophil-mediated activities and contribute to increase the inflammatory stress.
- **NK cells** display a remarkable cytotoxic character, turned on when they detect via their PRRs unusual amounts of specialized molecules in the surface of infected host cells or tumoral cells, thus coercing them to suicide and neutralizing the thread. This mechanism is singularly fast and subsequent inflammation is not always needed.

As mentioned above, lymphocytes are subdivided in two subtypes: B cells and T cells.

- **B lymphocytes** circulate in the blood and lymph in an immature or resting state. Through their B cell receptors (BCRs), they detect in a very specific fashion intact antigens either as soluble molecules or attached to pathogens, T cells or other antigen-presenting cells such as DCs. Once activated, they generate a progeny of plasma cells that produce and release soluble antibodies, intended for targeting and marking pathogens for clearance (humoral response).
- **T lymphocytes** lack the ability to recognize native antigens. Instead, their specialized T cell receptors (TCRs) are adapted to target a specific complex existing at the surface of dendritic cells and macrophages, called peptide-MHC complex. The peptide fragment belongs to a pathogen previously internalized and processed by the innate leucocyte. The MHC counterpart can be of two types: the MHC class I is recognized by cytotoxic T cells, also known as CD8 T cells since CD8 is the co-receptor

involved in such recognition process. As a result of the interaction, these lymphocytes induce cell apoptosis, similar to B cells. The MHC class II is recognized by helper T cells, analogously named CD4 T cells, in which the CD4 molecule works as the co-receptor. These immune cells differentiate under the influence of different cytokines to accomplish regulatory functions for innate and adaptive responses. As examples, T_h1 cells stimulate the cytotoxic activity mediated by CD8 T cells and B cells, whereas T_h2 and T_h17 lymphocytes recruit and maintain eosinophils and neutrophils, respectively. T_{reg} cells are essential lymphocytes specialized in suppressing the responses of other immune cells, contributing to maintain **tolerance to self-antigens**.

1.2. Dendritic cells and DC-SIGN: the biological context

DC-SIGN (Dendritic Cell-Specific ICAM-3-Grabbing Non-integrin) is one of the multiple pathogen recognition receptors (PRRs) that dendritic cells possess to recognize both foreign and self-antigens and to take the suitable actions to mediate intercellular communication and/or directly confront the danger. Amongst PRRs, DC-SIGN has awaken a particular interest for many scientists in diverse research areas due to its direct or indirect implication in a broad range of pathologies related to the immune system, including all kind of infections (virus, bacteria, fungi, parasites), autoimmune diseases and cancer. To uncover its role in so many different events, it is essential to firstly understand the biological and immunological context in which this receptor works, to finally address the quest for therapeutics from a structural point of view.

1.2.1. Dendritic cells

Dendritic cells, as well as macrophages, are usually referred to as **antigen presenting cells (APCs)** [9], given their efficient ability to process antigens taken from pathogens and present them to B and T lymphocytes. This route is crucial for the resultant adaptive response, since the specificity of individual T and B receptors is so high that an encounter with their specific antigens is rather unlikely. Dendritic cells can migrate to the lymph nodes and display notable amounts of MHC-peptide complexes, thus making possible the activation of hundreds to thousands of lymphocytes per single DC, all at once [10-11].

After their initial differentiation from stem cells, all dendritic cells remain in an immature stage (imDCs) and two “classical” types are clearly differentiated

[6,12]: the **interstitial or dermal DCs**, which stay inside the skin dermis, and **Langerhans cells (LCs)**, which finish their migration at the epidermis [13]. Both subtypes are involved in T cell activation, whereas B cell stimulation is exclusively supported by interstitial DCs [14]. Other subsets of DCs have been also described and studied:

- **Follicular dendritic cells (FDCs)** do not develop at the bone marrow and are involved in B cell proliferation along with interstitial DCs [15].
- **Plasmacytoid dendritic cells (pDCs).** Their roles remain still unclear, but they are thought to be involved in anti-viral defense through cytokine secretion [7].

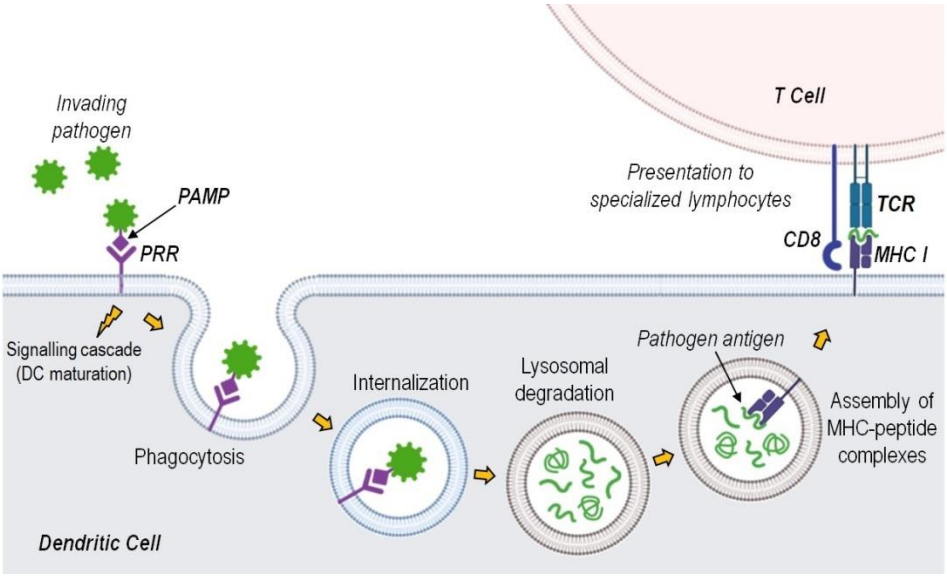


Figure 1.4. *DC response to an interaction with an external pathogen initiated by surface PRRs.*

Under their "guard" role, immature dendritic cells exhibit a great potential to detect and sample all kind of antigens by means of their PRRs, which chiefly include **C-type lectin-like receptors (CLR**) and **TOLL-like receptors (TLR).** After targeting a specific molecular pattern, the immature DC internalizes the antigen/pathogen and its maturation is then started [16-18] (Figure 1.4): the cell upregulates the production of co-stimulatory molecules and cytokines and drastically loss its ability to capture antigens by reducing the expression of PRRs [19-20]. Simultaneously, the DC activates the corresponding cellular machinery to process the pathogen components and assemble the resulting peptides to the MHC molecule [21-22]. Finally, developing dendritic cells

migrate to the nearby lymph nodes [23] to present the peptide-loaded MHC complex at their surfaces for T cell inspection, and eventually die.

Finally, it is worth noting the role of certain DCs within the context of immunological tolerance. To avoid lymphocytes recognizing self-antigens and attacking the host cells, the immune system rely on two strategies to ensure tolerance against self-antigens [24]. Some DCs in the medulla contribute to delete maturing self-reactive T cells (central tolerance) [25]. **Lymphoid DCs** are long-lived cells that may make use of the previously mentioned routes to present self-antigens attached to MHC molecules and silence those T cell populations that react to self-motifs (peripheral tolerance) [26].

1.2.2. C-type lectins and DC-SIGN in immunity

The functional differentiation between DC subsets is underlined by the presence of many different pathogen recognition receptors. Some of these PRRs are fairly specific for certain infectious agents, namely TOLL-like, NOD-like and RIG-like receptors [27-29]. However, other molecular sensors like C-type lectins are typical examples of receptors with a broad pattern of recognition [30]. Indeed, specificities between lectins usually overlap with each other [31-32] (Figure 1.5). Then, DC specialization depends on their PRR palette, the type of response that each lectin triggers according to the antigen that is recognizing, and the interplay between different PRRs operating at once [33]:

- Different lectin distribution: for instance, the macrophage mannose receptor (**MMR**) is highly expressed on macrophages [34-35]. **DC-SIGN** and the macrophage galactose-type lectin (**MGL**) are extensively expressed on interstitial DCs, although DC-SIGN is present in DCs from lymph nodes and mucosae as well [36], while MGL is preferentially found in those DCs maintaining immunological tolerance [37]. Finally, **Langerin** production is limited to Langerhans cells [38].
- Type of response: some receptors can lead DCs to stimulate inflammatory responses. For example, **Dectin-1** initiates inflammatory processes after binding to β-glucans [39], **DC-SIGN** after targeting Lewis antigens on *H. pylori* lipopolysaccharides [40], and **TOLL-like receptors** are key promoters of many inflammatory pathways against different foreign organisms [41]. Conversely, other C-type lectins participate in regulatory processes aimed at maintaining immunological tolerance and establishing cellular cross-talk. As examples, **MGL** is involved in T cell apoptosis through recognition of terminal α–GalNAc residues [42], the dendritic cell

immunoreceptor (**DCIR**) acts as mediator for neutrophil repression [43], and again, **DC-SIGN** is a crucial anchor for rolling of immature DCs as well as for DC-T cell synapse formation, interactions respectively established with fucose-containing ICAM-2 and ICAM-3 glycoproteins [44-45].

- Simultaneous triggering of several PRRs: it is common that certain responses are orchestrated by the joint action of C-type lectins and TOLL-like receptors. Namely, **Dectin-1** can conduct the production of cytokines in synergy with TLR-2 [46]. Also, some responses initiated by **DC-SIGN** combine with downstream TLR signaling to upregulate cytokine production [47].

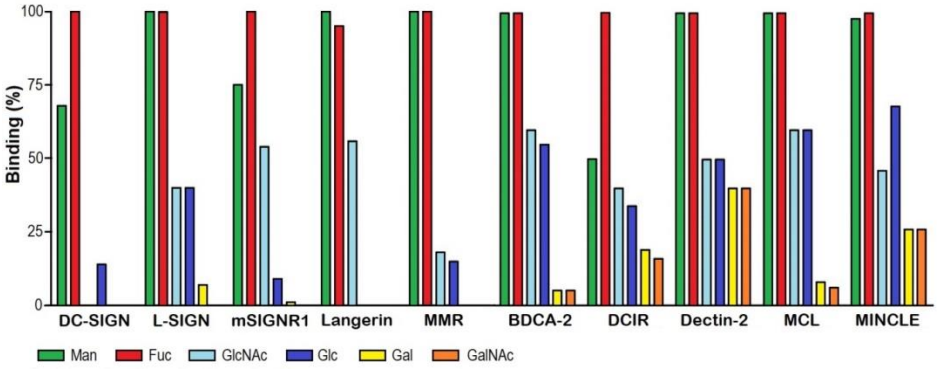


Figure 1.5. *Ligand-binding assays performed for six typical monosaccharides conjugated to BSA, using 10 common lectins (all human except for SIGNR1) as receptors. As noted, all of them can simultaneously recognize D-mannose (Man) and L-fucose (Fuc), and some of them, like BDCA-2 or MCL, have a broader selectivity, remarkably recognizing at least four hexoses. Adapted from [32].*

Taken together, the combination of all these factors will decide the ultimate outcome of an immune response. To note, PRRs are not mere receptors, but the antigen recognition also leads to signal transfer to the cytoplasm via different signaling pathways [48-49]. These cascades strongly depend on the specific recognized antigen and, also important, its three-dimensional presentation. For most of C-type lectins, these antigenic structures are complex **oligosaccharides**. DC-SIGN is not an exception [50], and its binding preferences will be extensively addressed in the next section from a structural point of view. Briefly, DC-SIGN essentially interacts with glycans containing **D-mannose** (Man) and **L-fucose** (Fuc) [51]. Notably, the DC-SIGN-mediated cytoplasmic signaling will be different depending on which sugar is recognized, giving rise to disparate immune responses [52-53] (Figure 1.6):

- Recognition of Man, for example in surface glycans from *C. albicans* [54] and *HIV* [55], helps DC-SIGN to modulate a pro-inflammatory cytokine programme through RAF1 activation. DC-SIGN cannot directly trigger the release of cytokines, but it modulates the participation of other receptors like TLR2 and TLR4 [56].
- Recognition of Fuc may have a dual role in immunity: on the one hand, DC-SIGN interaction with Fuc-containing PAMPs, like those from *S. mansoni* [57], leads to the activation of the Fuc signalosome, dissimilar to RAF1, which eventually promotes a long-term humoral immunity against the pathogen [58]. On the other hand, DC-SIGN is thought to be involved in sustaining immune tolerance through recognition of self-glycoproteins with antigens bearing Fuc, such as **Lewis and blood group antigens**. Alike, DC-SIGN take part in cell-cell communication, for instance through interaction with Le^Y antigens of ICAM-2 (DC adhesion and rolling) [59] and the macrophage receptor 1, Mac-1 (DC-neutrophil clustering) [60].

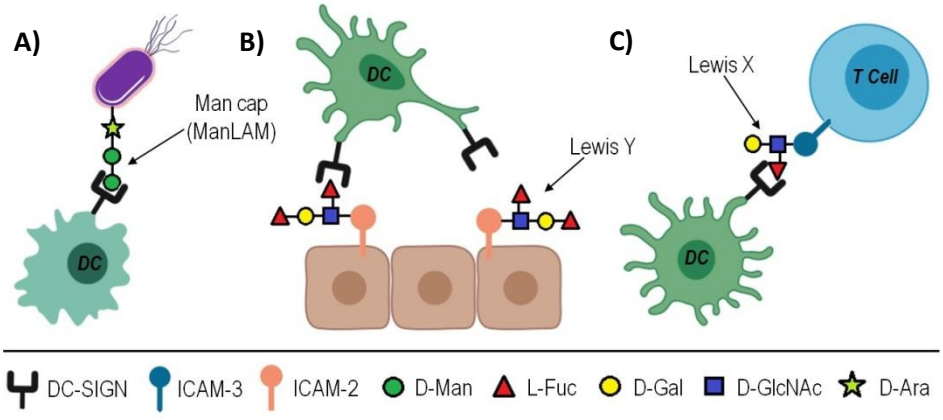


Figure 1.6. *DC functions mediated by glycan recognition through the DC-SIGN receptor. A) DC maturation and secretion of pro-inflammatory cytokines after recognizing pathogenic Man-containing antigens. B) DC adhesion and rolling based on recognition of Le^Y motifs. C) DC-T cell crosstalk established by interaction with the Le^X antigen.*

1.2.3. When the immune system fails: DC-SIGN as Trojan horse for disease development

As DC-SIGN, some C-type lectin receptors lack the ability to autonomously trigger inflammation, whereas other CLRs are "self-sufficient" for that task. The abilities of the different PRRs to recognize invading organisms through **carbohydrate-mediated interactions** have been thoroughly studied over the

last decades [61]. In most of cases, these receptors predominantly take up beneficial pro-inflammatory responses that aid the surrounding tissues to repair damages or oppose to an infection. Thus, **Dectin-1** and **Dectin-2** are essential to repel fungal infections by *C. albicans* [62] and *Aspergillus* spp. [63], respectively, while **MGL** increases macrophage activation to kill parasites like *T. cruzi.* [64]. The macrophage C-type lectin (**MCL**) triggers protective mechanisms against *M. tuberculosis* [65], and **Mincle** does it against *S. pneumoniae* [66] and *K. pneumoniae* [67]. **Langerin** interacts with HIV and has been found to efficiently promote viral clearance at Birbeck granules, avoiding T cell infection [68]. Also, MCL and Dectin-2 have been described to help in clearance of tumor cells [69].

But unfortunately, the immune system is not infallible at all: the same mechanisms may otherwise contribute to worsen the pathological state [70-71]. The quick evolutionary skills of many pathogens bring about overwhelming situations which the immune system cannot cope with. Thus, some pathogens may be able to subvert the effector pathways to ensure its entry and dissemination, chiefly by taking advantage of lectin recognition and attachment [72-73]. In this context, **DC-SIGN** is actually a paradigmatic case, as this receptor is greatly used by numerous pathogens for their own benefit [74]. Broadly speaking, many microorganisms gain access to adaptive T helper cells by using DC-SIGN as a vehicle to modulate DC activity [75-76], thereby easing disease persistence:

- DC-SIGN-mediated ***human immunodeficiency virus*** **(HIV)** transmission was the first reported case of an infection in which a C-type lectin played a crucial role [77]. Once internalized, the viral particles may survive for prolonged periods by biasing the endocytic routes of the dendritic cell [78] and/or increasing DC-SIGN expression to boost the attachment of more viral particles to the membrane, until final T cell targeting [79-80]. Simultaneous TLR8 and DC-SIGN targeting facilitates the infection of DCs themselves as well [81]. In all cases, the virus-host union is established through recognition of the viral glycoprotein gp120 (Figure 1.7A).
- The ability of ***dengue virus*** **(DENV)** to hold on to the cell membrane is entirely ascribed to DC-SIGN interacting with the viral E-envelope glycoprotein [82-83]. Indeed, the carbohydrate-lectin interface has been already described by using cryoEM [84]. However, a marked controversy currently surrounds the exact role of DC-SIGN in this viral infection. DC-SIGN has been newly suggested to mediate viral entry as well as pathogen anchoring, without intervention from other membrane co-

receptors [85]. In contrast, previous works postulated that DC-SIGN is dispensable for viral entry [86-87].

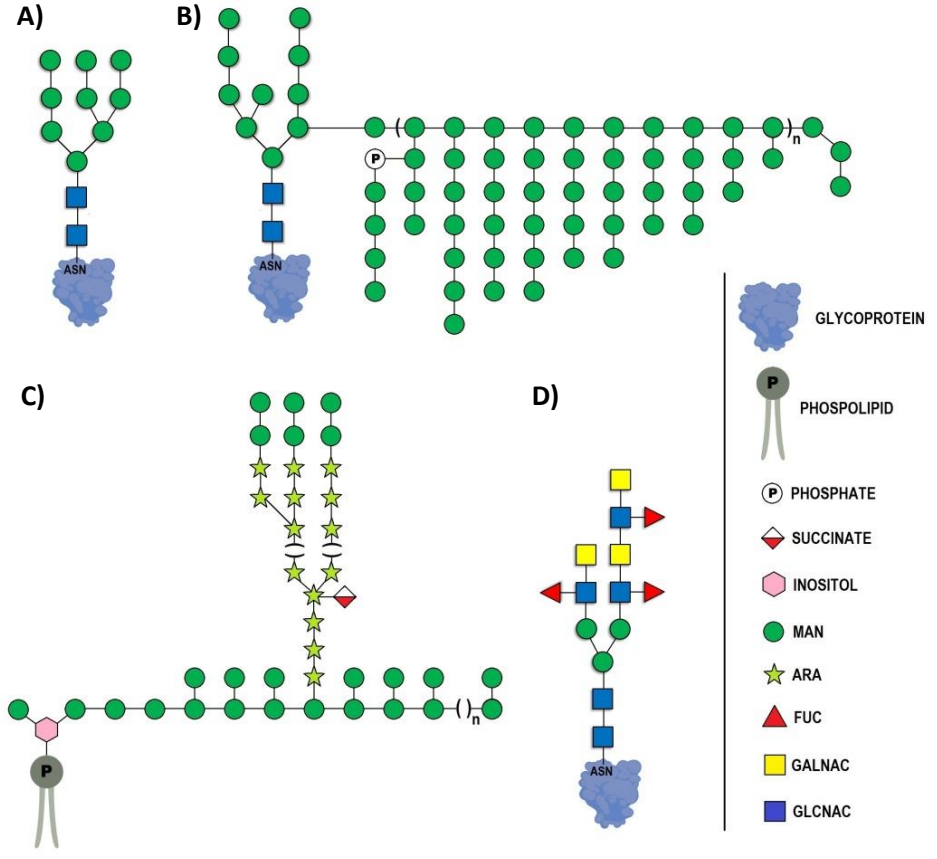


Figure 1.7. *Typical carbohydrate antigens from pathogens that are recognized by DC-SIGN. A) Viral high-mannose N-glycan (VIH). B) Fungal complex N-mannan (C. Albicans). C) Bacterial Man-capped LAM (M. Tuberculosis). D) Representative Fuc-containing glycoprotein from S. mansoni´s SEAs.*

- Infections in DCs driven by ***Ebola virus*** **(EBOV)** have been reported to involve both DC-SIGN and its close homologue, L-SIGN [88]. Both lectins interact with mannosylated glycoproteins on the viral envelope (GP), triggering similar evasion pathways to reach T cells as those exerted by HIV [89]. Otherwise, DC-SIGN merely acts as a crucial anchor for the pathogen, but do not guide viral internalization. In this regard, other research lines have been focused on identifying such entry routes [90-92].
- Regarding other viral-related pathologies, binding of ***human cytomegalovirus*** **(hCMV)** [93-94] and ***hepatitis C virus*** [95-96] to DC-SIGN was already documented the past decade, and more recently, this C-type

receptor has been postulated to play analogous roles in several other diseases, like those originated by ***Japanese encephalitis virus*** [97], ***Rift valley fever virus*** [98], ***Kaposi´s sarcoma associated herpesvirus*** **(KSHV)** [99], ***measles virus*** [100] and ***phlebovirus*** [101]. Interestingly, DC-SIGN and L-SIGN may take part in flu episodes caused by ***influenza A virus*** **(IAV)**, but as co-receptors for viral entry [102] rather than attachment proteins (this first step is mediated by binding of the influenza hemagglutinin to sialic acid moieties).

- Moving to bacterial infections***,*** DC-SIGN has been certainly linked to the deadly propagation routes of ***Mycobacterium tuberculosis*** in host cells [103-104]. This disease is one of the most significant worldwide health risks and has encouraged many researchers to focus on DC-SIGN as a therapeutic target. As opposed to HIV, *M. tuberculosis*, via Man-capped lipoarabinomannan (ManLAM, Figure 1.7C), binds to DC-SIGN to only suppress DC maturation signaling [105], precluding effector mechanisms in the surrounding environment such as autophagy or apoptosis [106], and consequently escaping from the adaptive response and ensuring its own surveillance [107-108]. In the study of ***Porphyromonas gingivalis***, a similar subversive route involving DC-SIGN-TLR2 cross-talk has been identified as well [109].
- Fungal infections, especially those involving ***Candida*** species, have also attracted attention on DC-SIGN [54]. Notably, many C-type lectins actually contribute to recognize the PAMPs on the cell wall [110-111], although DC-SIGN and MMR seem to be the main mediators in DCs. Specifically, DC-SIGN bind to complex highly-mannosylated N-glycans on yeast´s cell surface [112] (Figure 1.7B). The same glycans are exploited by ***Saccharomyces*** [113] and ***Paraccidioides*** species [114] to target DCs via this protein receptor.
- Finally, DC-SIGN has proven to be relevant in the persistence of parasitic infestations, being those caused by ***S. mansoni*** [115] and ***L. pifanoi*** [116] the most explored ones. In fact, the former case is an illustrative example of DC-SIGN targeting driven by Fuc-containing antigens, essentially Lewis-type structures on the soluble egg antigen (SEA) of *S. mansoni* (Figure 1.7D). By contrast, the carbohydrate antigens in *Leishmania* amastigotes remain unclear yet [117].

In the same line, tumor cells also spread out and suppress immune responses by means of sugar-mediated interactions, originating cancer. **MGL** contributes to this cause through interaction with tumor-related glycoforms of MUC1 [118-119]. Analogously, **DC-SIGN** impairs DC maturation when targeting blood

group antigens from tumor-derived CEACAM-1 glycoprotein [120]. In other cases, the regulatory mechanisms fail and the inflammatory response is not well controlled, thereby exerting harmful effects on the nearby regions [121]. In absence of any infection, this malfunctioning gives rise to autoimmune diseases. The direct implication of C-type lectins in these disorders has been guessed: **MBL** expression has been considered as a risk factor in diabetic complications [122], while **DCIR2** expression may be relevant in multiple sclerosis [123]. **DC-SIGN**, as well as **MMR**, have been linked to exacerbated allergic responses triggered by allergens in peanut and pollen [124-125].

1.3. C-type lectins and DC-SIGN: structural insights

Historically, the term "*C-type lectin*" was coined to refer to all these proteins which interact with carbohydrates in a calcium-dependent manner, through a compact globular structure called "*carbohydrate recognition domain*" (**CRD**) [126]. Since the late eighties, this family has steadily grown and currently includes more than 1000 protein receptors classified into seventeen subgroups (Figure 1.8). However, the term "lectin" would not be strictly correct to designate this family, since that "CRD" have been found to be a well-conserved structural scaffold shared by many other receptors that, in some cases, do not bind to carbohydrates or even to calcium [127-128]. Instead, the term "*C-type lectin-like domain*" (**CTLD**) is less ambiguous and is often used to include all subgroups [129]. Even so, the term "*C-type lectin*" is still used in a wide scope, including members that recognize other substrates like proteins or lipids, whereas the term "*CRD*" is obviously reserved for only referencing lectin domains with carbohydrate-binding ability.

From a structural perspective, all C-type lectin-like domains share a series of common structural elements that are highly conserved among different species [130-131].

The CTLD fold consists of a central core essentially constituted by five β-strands flanked by two α-helices that lie in perpendicular to each other (Figure 1.9). The first and last β-strands (β_1 and β_5) form an antiparallel β-sheet on the bottom part of the structure, where both the N- and the C-terminus are located. The three remaining β-strands (β_2, β_3 and β_4) are assembled into a β-sheet that constitutes a quite shallow plateau rather exposed to the external environment. Finally, canonical CTLDs also exhibit a characteristic and long loop region comprising about 30 amino acids, that extends from the end of β_2, going around the area limited by the three-stranded β-sheet and the α_2-helix.

In carbohydrate-binding C-type lectins, this loop takes part in glycan recognition.

Apart, the β_2-strand typically possesses a highly conserved "WIGL" motif that is often used as a landmark for sequence analysis [132-134]. There are also four conserved cysteine residues that establish two disulfide bridges between α_1-β_5 and β_3-β_4, respectively (Figure 1.9).

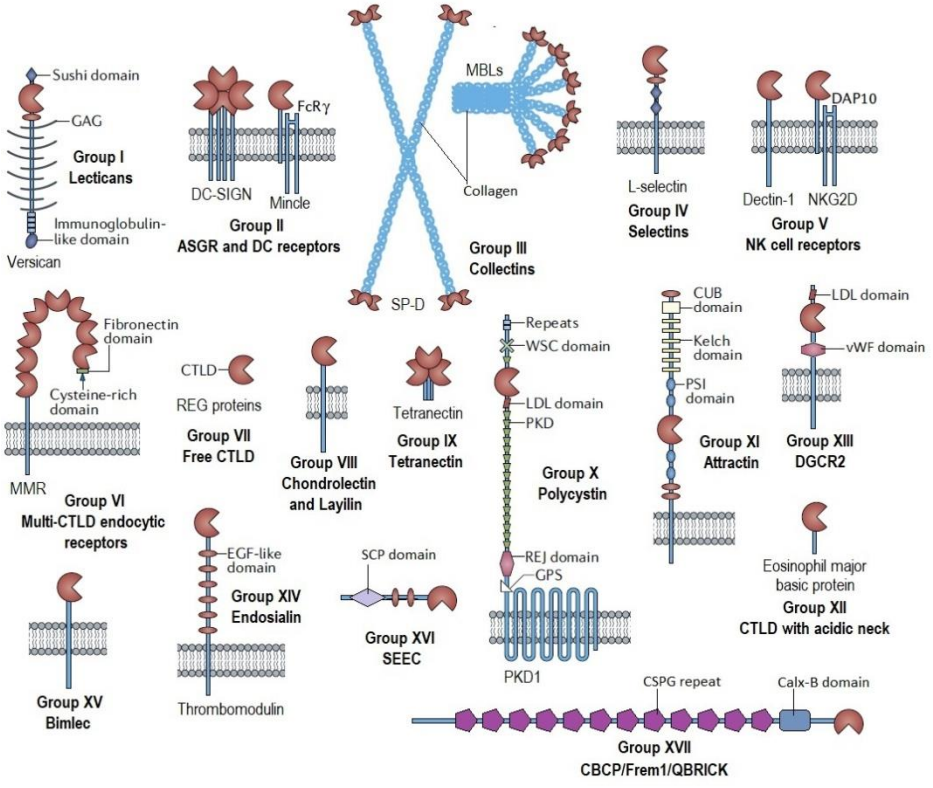


Figure 1.8. *Cartoon representation of the seventeen subgroups that constitute the C-type lectin superfamily (Adapted from [130] and [61]). Calx-B: calcium binding domain. CSPG: chondroitin sulphate proteoglycan. CUB: complement C1r/C1s, UEGF, BMP1. EGF: epidermal growth factor. FcRγ: Fc receptor γ-chain. GAG: glycosaminoglycan. GPS: G protein-coupled receptor proteolytic site. LDL: low-density lipoprotein. PKD: polycystic kidney disease. PSI: plexin–semaphorin–integrin. REJ: receptor for egg jelly. SCP: sperm coating glycoprotein. SP-D: surfactant protein D. vWF: von Willebrand factor. WSC: cell wall integrity and stress response components.*

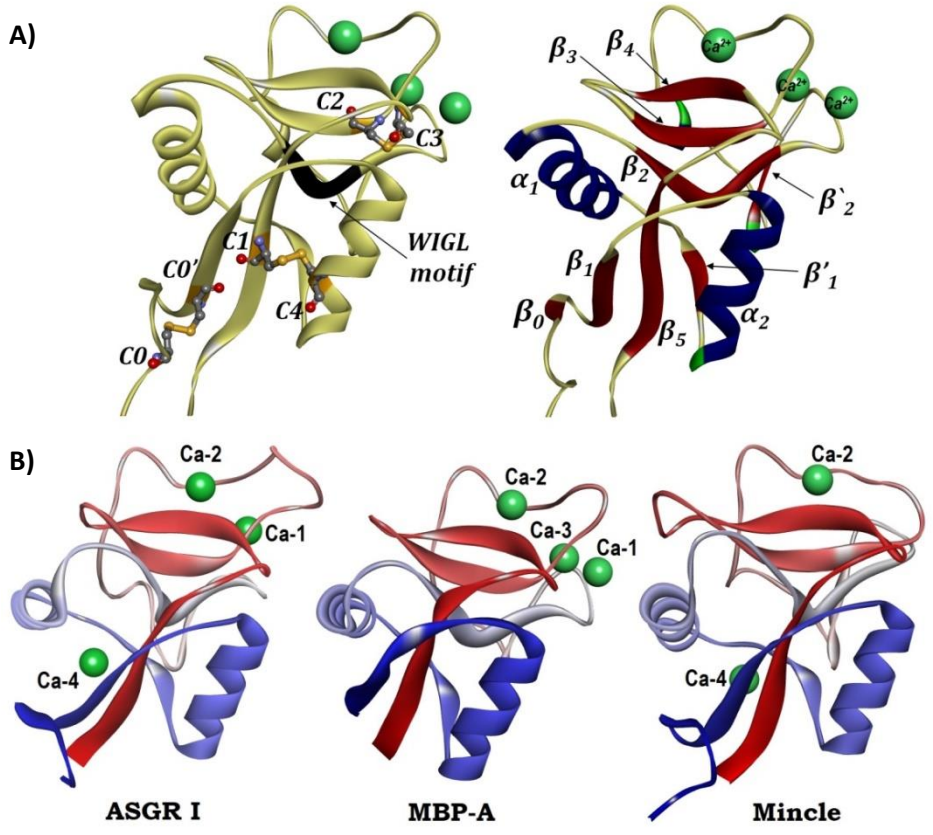


Figure 1.9. *A) On the left, the DC-SIGN CRD (PDB: 1SL5) displaying the characteristic structural signatures like the WIGL motif (black ribbon) and the four cysteines (C1-C4) that form disulfide bridges. On the right, the typical secondary elements that constitute the CTLD fold. B) Comparison between the CRDs of three representative lectins from subgroup II which show different occupied calcium sites (ions represented as green spheres). PDBs (from left): 1DV8, 2MSB and 3WH2.*

Regarding the calcium ions, at least four Ca^{2+}-binding sites have been observed so far, although not all of them are necessarily occupied [135-136]. Certainly, crystallographic structures have shown lectins with one, two, three or even no calcium ions. All calcium atoms, if present, play roles in folding stability and structural integrity, but only Ca^{2+} ion at site 2 is directly involved in carbohydrate binding [31,137]. Sites 1 and 3 contribute to improve the stability of the long loop region [138-140], as observed for DC-SIGN [141]. Calcium at site 4, as reported for the asialoglycoprotein receptor I (ASGR-I) [142], establishes salt bridges between the interface of the β_3-β_4 sheet and the neighboring α_2-helix.

Calcium at site 2 is particularly important due to its biological relevance in **lectin-like activity**. In fact, conformational changes derived from the presence or absence of Ca^{2+} sometimes have biological significance [143]. Thus, C-type lectins like tetranectin, which does not even bind to carbohydrates, undergo structural rearrangements in the long loop region after losing Ca^{2+} at site 2, enabling the binding to plasminogen [144]. Similarly, pH-induced destabilization of the loop regions, followed by calcium loss, leads those carbohydrate-binding C-type lectins acting as endocytic receptors to release the cargo once internalized [145-146].

Structurally, binding of Ca^{2+} at site 2 is sustained by three consecutive residues of the long loop, including a central proline in a *cis* conformation, other three residues on the nearby β_4-strand, which form the so-called "WND" motif, and an additional residue located close to the beginning of the loop [143]. The *cis*-proline and the tryptophan from the WND motif do not participate in calcium binding, but are highly conserved among all CTLDs and therefore considered [132]. The other cited residues establish coordination bonds with the Ca^{2+} through their carbonyl groups. The 7- or 8-coordinated sphere is completed with water molecules in the unbound form [147], and with two hydroxyl groups from a single monosaccharide in the sugar-protein complex. Interestingly, earliest observations already pointed out that the sugar specificity depends on the specific residues flanking the *cis*-proline [148]:

- The C-type lectins displaying the **EPN motif** are known as mannose-binding proteins, as they often manifest a marked binding preference for sugars with equatorial hydroxyl groups at positions C3 and C4, mainly D-Man and D-glucose (Glc). This group includes C-type lectins like **DC-SIGN, Langerin** [149] and **LSECtin** [150].
- The C-type lectins containing the **QPD motif** are commonly named galactose-binding proteins. These receptors better bind to sugars with an axial hydroxyl group at C4, like D-galactose (Gal). The macrophage Gal-specific lectin (**MGL**) [151-152] is the most representative member of this group, which also includes the asialoglycoprotein receptor (**ASGR**) [153] and the scavenger receptor C-type lectin (**SRCL**) [154-155].

Overall, the binding site region is constituted by the portion of the long loop surrounding the Ca^{2+} at site 2 and the adjacent area occupied by the three-stranded β-sheet. As mentioned above, the interaction with the sugar is essentially established by coordination of two adjacent hydroxyls groups to the calcium ion. Considering that the binding area is fairly shallow and solvent-

exposed in most of cases, this means that such an interaction could be easily disrupted in an aqueous environment and perfectly explains the **low intrinsic affinities** often found in sugar-binding C-type lectins (Figure 1.10).

	DC-SIGN[a]	DC-SIGNR[a]	Langerin[b]	MBP-A[c]	MBP-C[c]
Man	13.1	2.6	2.3	-	-
Gal	72	20	16.1	212	114
Glc	23	7.3	7.0	-	-
Fuc	6.7	4.1	2.6	2.9*	8.8*
α-OMe-Man	12.5	3.4	1.7	8.0	5.7
α-OMe-Gal	270	112	>200	-	-
α-OMe-Glc	32	9.1	-	-	-
ManNAc	8.7	1.3	5.6	-	-
GlcNAc	32	4.7	2.8	12.7*	14.4*

Figure 1.10. *Dissociation constants (in mM) determined for several monosaccharides by means of competitive assays, using five lectins including DC-SIGN. [a]) Ref. [156]. [b]) Ref. [149]. [c]) Ref. [157]. *) Data only available for the Methyl-α-glycosides.*

Secondly, the sugar-lectin complex is stabilized by an additional network of hydrogen bonds primarily involving the same coordinated OH groups and the surrounding polar sidechains [158-160]. However, this mode of interaction, especially in receptors such as DC-SIGN, is geometrically rather "malleable" as long as the positioning of hydrogen donors and acceptors does not originate steric clashes or electrostatic repulsions. In fact, for Man-binding proteins (MBPs), the interaction has been proved to take place in two orientations: the conventional mode and an alternative mode where the pyranose ring of the sugar is rotated 180º [157,161]. Even the EPN/QPD division neither satisfies the classification criteria, as for instance, several Man-binding receptors (like selectins or Langerin) are known to recognize Fuc as well [162-163], in which OH4 display an axial configuration. Hence, it may occur that several different sugars fulfill the geometrical requirements to target the binding site, giving rise to another characteristic feature of C-type lectins: **broad ligand specificity** (Figure 1.10).

Moreover, it becomes clear that binding specificity cannot be limited to monosaccharide recognition, as C-type lectins are normally exposed to complex oligosaccharides in the biological milieu. Aside from direct polar contacts, van der Waals interactions arising from the surrounding monosaccharides also contribute to favor or preclude sugar attachment [164-165], depending on the particular three-dimensional geometry and the existing differences in the exposed amino acids for each lectin.

Finally, the poor affinities agree with the observation that CTLDs usually oligomerize into multimeric ensembles (dimers [166-167], trimers [168-169] or tetramers [170-171]), thereby increasing their ability to target glycans by means of more complex **multivalent interactions**. The important features underlying these interactions will be subsequently detailed for DC-SIGN.

1.3.1. The structure of DC-SIGN

DC-SIGN belongs to the group II of C-type lectins, one of the largest and most heterogeneous groups (Figure 1.11). All proteins inside this classification are transmembrane receptors with a common structural organization: a short cytoplasmic tail, a transmembrane domain and an extracellular domain [130]. The latter one consists of the carbohydrate recognition domain (CRD), and a stalk or neck region that is involved in protein oligomerization and anchors the CRD to the membrane (Figure 1.12). The length of this neck is widely variable among different members.

In particular, the extracellular domain (**ECD**) of DC-SIGN (CRD + neck) is tetrameric. The neck region comprises 176 amino acids grouped into seven highly conserved 23-residue repeats, and an additional truncated 15-residue repeat at the N-terminus [172]. CRD and neck are structurally independent from each other, in contrast to other lectins like collectins, with a more packed architecture [173-174]. In fact, DC-SIGN tetramerization is also CRD-independent and crystallographic models have suggested that the CRD is flexibly linked to the neck, thus enabling local CRD motions and increasing the ability of DC-SIGN ECD to target complex branched oligosaccharides in a multivalent fashion [175].

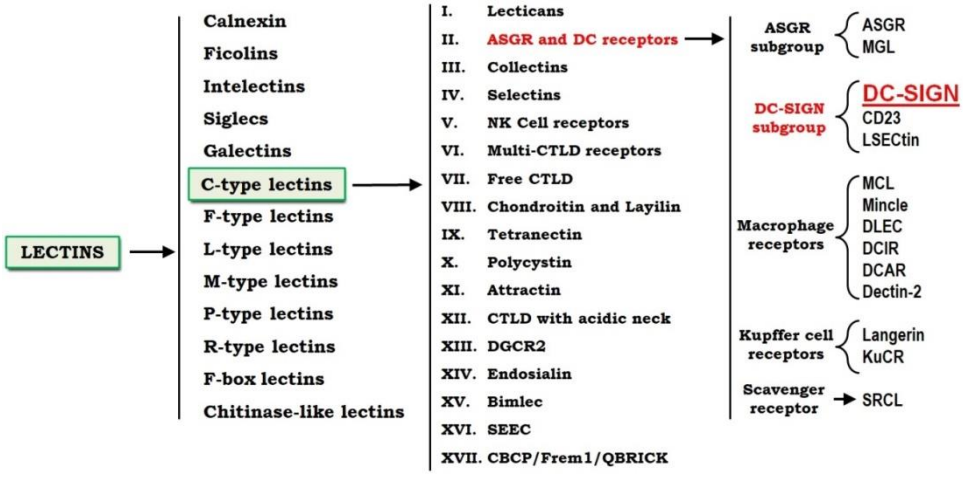


Figure 1.11. *Classification of DC-SIGN within the global denomination of "lectin" [176].*

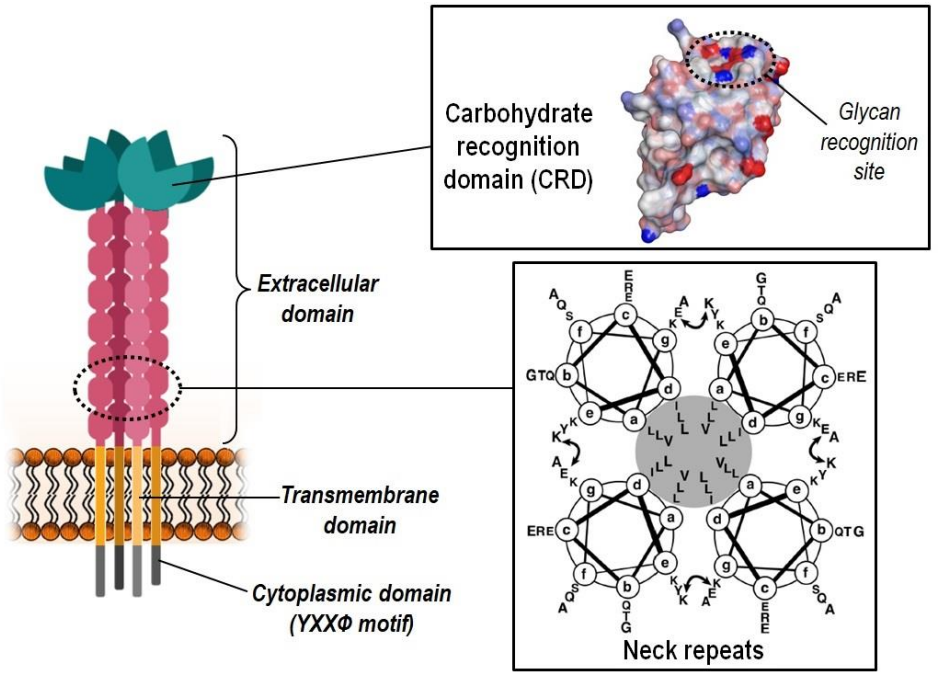


Figure 1.12. *Schematic representation of the oligomeric ensemble that DC-SIGN adopts at the cell membrane. The top box displays the globular structure of the CRDs. The bottom box shows a cross view of the four neck domains associated through hydrophobic contacts from the stalk core (highlighted in grey) and salt bridges between external Glu and Lys sidechains (indicated with arrows). Adapted from [177]. The short cytoplasmic domain consists of an YXXΦ motif [178].*

Each 23-amino acid repeat forms an α-helix and the four necks tightly pack between them constituting a robust bundle primarily stabilized by hydrophobic interactions arising from the hydrophobic residues buried into the stalk core. Otherwise, the existence of the tetrameric ensemble requires at least the presence of six amino-acid repeats (78% of the neck sequence). Removal of more repeats leads to equilibrium between tetramers and dimers, and with two repeats or less the tetramer is no longer stable and only dimers and monomers are detectable [179]. Subsequent studies revealed that the ECD stability might not exclusively depend on the hydrophobic core, as the ECD undergoes a deep loss of its secondary structure under acidic conditions. Low pH values not only preclude ligand binding at the CRD, but also promote the loss of the helical structure of the neck, eventually causing the tetramer destabilization. Then, salt bridges between exposed Glu and Lys residues have been postulated to maintain the bundle assembly, and hence Glu re-protonation at acidic pHs would explain the loss of stability [177]. However,

this explanation is still controversial, as the tetramer remains totally stable at high ionic strengths.

Through diffusion and sedimentation coefficients, in combination with modeling, several models have been proposed for the ECD architecture (Figure 1.13). Apart from the parallel packing of all the α-helices, a different Y-shaped structure has been proposed, with the last two repeats associated in antiparallel hairpins. The parallel distribution allows CRD flapping between closed and open forms, although SAXS analyses have shown that DC-SIGN might tend to maintain a more packed, closed CRD distribution.

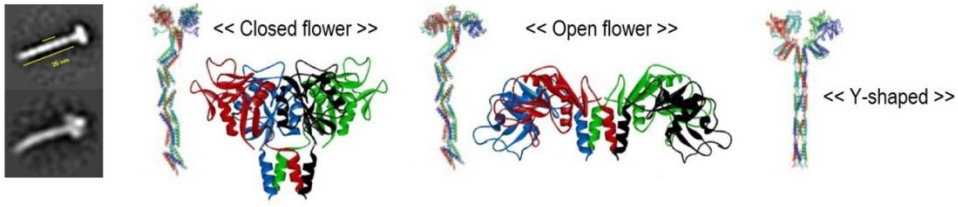


Figure 1.13. *The three tridimensional models of the DC-SIGN ECD that better fulfill the experimental data from sedimentation and diffusion studies. Adapted from [177,179]. On the left, a TEM image of the ECD used in this work (See Chapter 4.1).*

At the cell membrane, DC-SIGN tetramers freely diffuse throughout the cell surface and randomly cluster forming heterogeneous nanosized domains with different protein densities [180]. The formation of these nanodomains is driven by intermolecular interactions between the four-helix stalks. This nanoclustering could be aimed at maximizing the docking of DC-SIGN to big pathogens on larger scales [181].

1.4. DC-SIGN and carbohydrate recognition

DC-SIGN is the paradigmatic case of receptor binding promiscuity. Its broad specificity has been described for years. Akin to other C-type lectins, DC-SIGN specifically recognizes a plethora of simple monosaccharides with affinities in the mM range (Figure 1.10). D-Man is the best binder among D-sugars (K_i ≈ 13 mM), as expected from a lectin with the EPN sequence motif. In addition, the axial stereochemistry of the OH2 group seems to favor binding in N-acetyl derivatives, whereas the equatorial disposition is nearly irrelevant, as seen in Glc and derivatives (K_i ≈ 20-30 mM). On the contrary, the interaction with L-Fuc (OH2 is equatorial) is slightly better than D-Man (K_i ≈ 7 mM). Gal binding is extremely weak (20-fold worse than Man) and this sugar was actually considered as a non-binder [51,156].

As mentioned in the previous section, reaching a global understanding about DC-SIGN specificities towards complex oligosaccharides must not be limited to the simple sugar-mediated Ca^{2+} binding, but a thorough analysis of many other factors should be carried out, namely:

- Three-dimensional geometry of oligosaccharides: large glycans exhibit a great flexibility, and hence, a broad conformational space [182-183]. The frequencies of the different conformers may influence the accessibility of certain residues to the binding site, either due to possible steric clashes arising from nearby branches or to thermodynamic aspects like desolvation or loss of flexibility (enthalpy/entropy balance) [184-187].
- Multivalence mechanisms: on the one hand, large branched glycans typically contain several monosaccharides that can be independently recognized at the binding site. This gives rise to the so-called *"statistical rebinding"* effect, such that all possible modes of interaction synergistically improve the global binding constant [188-189]. On the other hand, complex oligosaccharides, if large enough, are able to simultaneously target several near CRDs originating "*chelating or clustering effects"* [190-192], which raise the affinity as well (Figure 1.14).

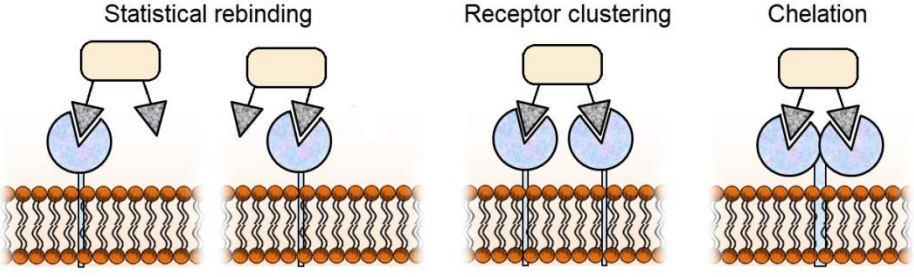


Figure 1.14. *Representative multivalence mechanisms involving complex glycans with multiple epitopes.*

- Secondary interactions on the protein surface: the suitable binding to the primary site is strongly influenced by the position of the interacting monosaccharide in the oligomeric structure. Namely, the glycosidic linkages define the spatial arrangement of the immediately linked sugars, determining which sugars could correctly accommodate at the binding site [162,193] (Figure 1.15).

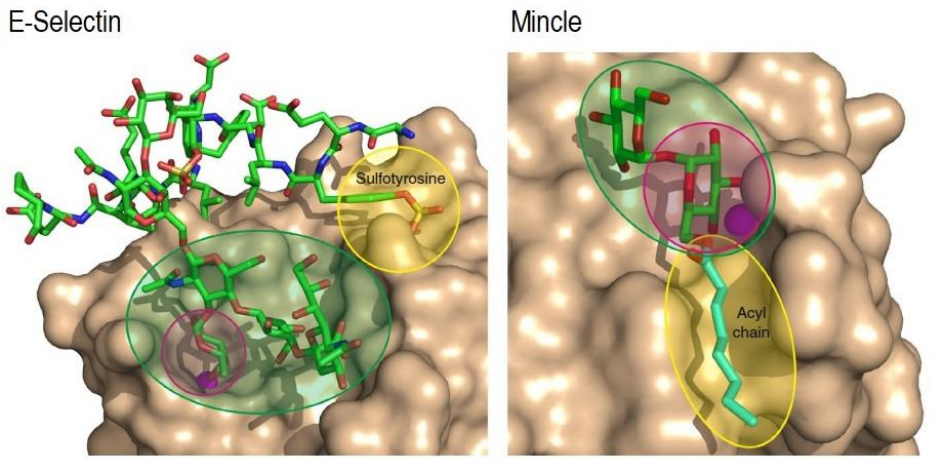


Figure 1.15. *Two examples of carbohydrate-lectin recognition events in which the primary sugar-lectin interaction (hexose circled in purple) is assisted by favorable secondary contacts arising from the surrounding sugar moieties (circled in green) or even from remote ligand groups that target other distal protein pockets (highlighted in yellow). On the left, E-selectin interacting with a peptidic chain containing the sialyl-Le^X epitope (PDB: 1G1S). On the right, mincle recognizing a glycolipid bearing a trehalose disaccharide (PDB: 4KZV). Adapted from [31].*

In particular, DC-SIGN has a fairly flat, solvent-exposed area around the calcium ion. This plateau is essentially composed of acidic amino acids (Asp, Glu) and several other non-charged or basic residues (Ser, Asn, Gln, Lys). Flanking this area, three protruding hydrophobic residues can be found: **Val351** at the long loop, quite near to the binding site and very important in fucose binding; **Phe313** at the edge of the three-stranded β-sheet, which may play stabilizing roles in the recognition of long Man-containing branches in mannans; and **Leu371**, less important and further away from the calcium site. This architecture is crucial to understand the different specificities between related lectins, for instance regarding **L-SIGN**, **Langerin** and **MBPs**, three C-type receptors usually studied alongside DC-SIGN (Figure 1.16):

- The CRD of **L-SIGN** (or DC-SIGNR) shares a lot of similarities with that of DC-SIGN [194]. However, the position equivalent to Val351 in L-SIGN is occupied by a serine residue (Ser363) [156]. This switch from an apolar residue to a polar one abrogates the binding to the Fuc-containing Lewis X motif, while Man recognition remains unaffected [195-196]. Other Lewis antigens, like Le^A or Le^B, still interact but their affinities are very poor.

- **Langerin** CRD shows two important differences in comparison with that of DC-SIGN: firstly, an alanine (Ala289) lies close to the binding site instead of Val351 [197-198]. Even so, the hydrophobic nature of the sidechain has not changed and Fuc binding is possible yet. Secondly, the binding site is less acidic and there are two protruding lysine sidechains (Lys299 and Lys313) closely flanking the metal ion, opposite to the loop. The presence of these residues has marked consequences: Fuc binding is limited (some Lewis-type antigens do not bind) [199]; the interaction with high-Man glycans is preferentially driven by the outer non-reducing residues, due to a smaller extended binding site compared with that of DC-SIGN [200]; and interestingly, the exposed amino groups can accommodate sulfated groups by means of a double salt bridge, opening the door to the recognition of sulfated sugars (like 6S-Gal), only described for this lectin so far [201-203].
- Mannose-binding proteins **(MBPs** or **MBLs)** belong to the group of collectins (Group III), but are another illustrative example of subtle structural modifications underlying notable differences in specificities. Many structural data about these lectins has been obtained from the rat homologues MBP-A and MBP-C**.** The CRD of **MBP-A** is more polar and devoid of relevant exposed hydrophobic residues [204]. It has a couple of protruding charged sidechains (Asp200, Lys182) which participate in binding to oligomannans, rather than hydrophobic amino acids. Indeed, Phe313 and Val351 positions on DC-SIGN are respectively substituted by a shorter serine sidechain and a bulky histidine residue close to the calcium ion (Ser154 and His189). The accommodation of mannans is thereby constrained by the presence of all these residues, and binding to long sugar chains preferentially takes place through the non-reducing end [205]. Instead, MBP-C has a valine residue and hence a binding behavior very similar to DC-SIGN [147]. Interestingly, not only MBP-C but also MBP-A can recognize Fuc, although the stabilizing role of His189 in the latter case is unclear yet [157].

Overall, all these factors should be taken into account for exploiting this receptor as a therapeutic target for drug development. In this line, the next and last sections summarize the current knowledge about DC-SIGN-mediated recognition of natural ligands, as well as related mimetics and other compounds that have successfully exploited secondary interactions and multivalency to achieve strong affinities and compete with PAMPs.

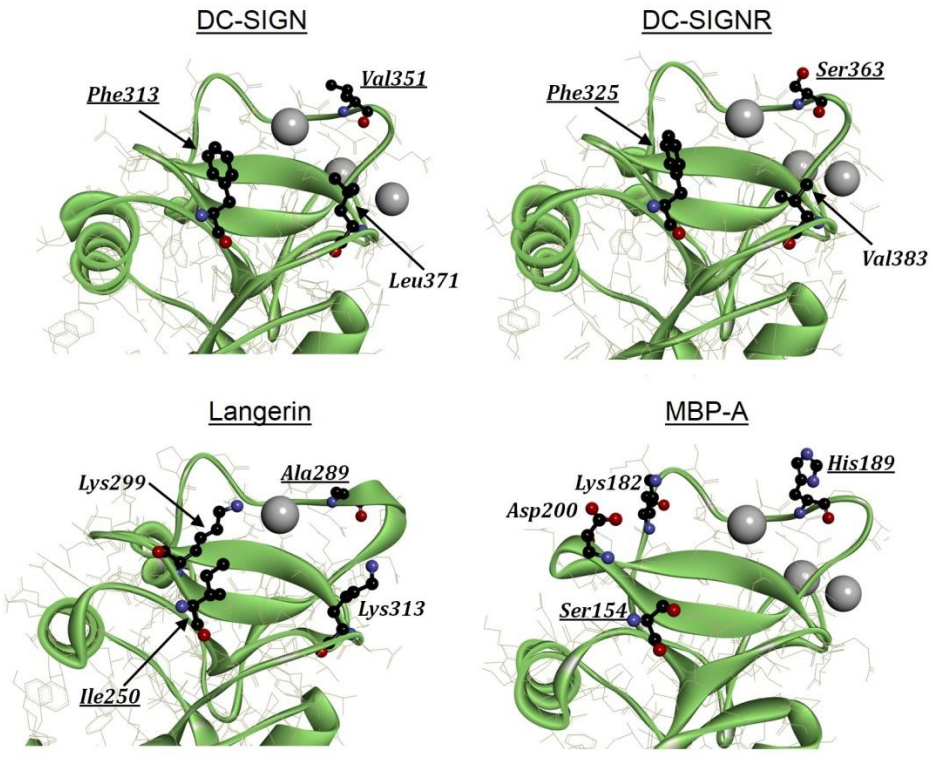


Figure 1.16. *Structural comparison between DC-SIGN and three closely related C-type lectins: DC-SIGNR (PDB: 1K9J), MBP-A (PDB: 2MSB) and Langerin (PDB: 3C22). Those sidechains that notably influence the glycan arrangement at the recognition site are detailed in each case. The underlined residues are those positions structurally equivalent to Phe313 and Val351 in DC-SIGN.*

1.4.1. Mannose-based binding

Given that most of pathogens interact with DC-SIGN through Man-containing antigens, the recognition of this sugar has been extensively studied from a structural point of view.

Early research immediately evidenced the structural versatility of the DC-SIGN CRD when binding to complex mannans. Hence, it is known to recognize both outer and inner Man residues linked by different glycosidic bonds while perfectly accommodating the surrounding mannoses by means of secondary intermolecular interactions. Starting from the architecture of the prototype mannan, Man_9, many studies have shed light into the structural preferences of DC-SIGN using fragments derived from the original Man_9 scaffold, thus reducing the complexity of the system.

From a broad perspective, DC-SIGN exhibits a marked preference for the reducing moiety of the Manα1-2Man disaccharide unit. This interaction is fairly favorable regardless the glycosidic linkage that attaches the disaccharide to the core structure (α1-2, α1-3 or α1-6). Additionally, DC-SIGN binds to both non-reducing mannoses of the core Man_3 trisaccharide (Manα1-3[Manα1-6]Man), although the binding to the α1-3-linked Man seems to be slightly favored [206].

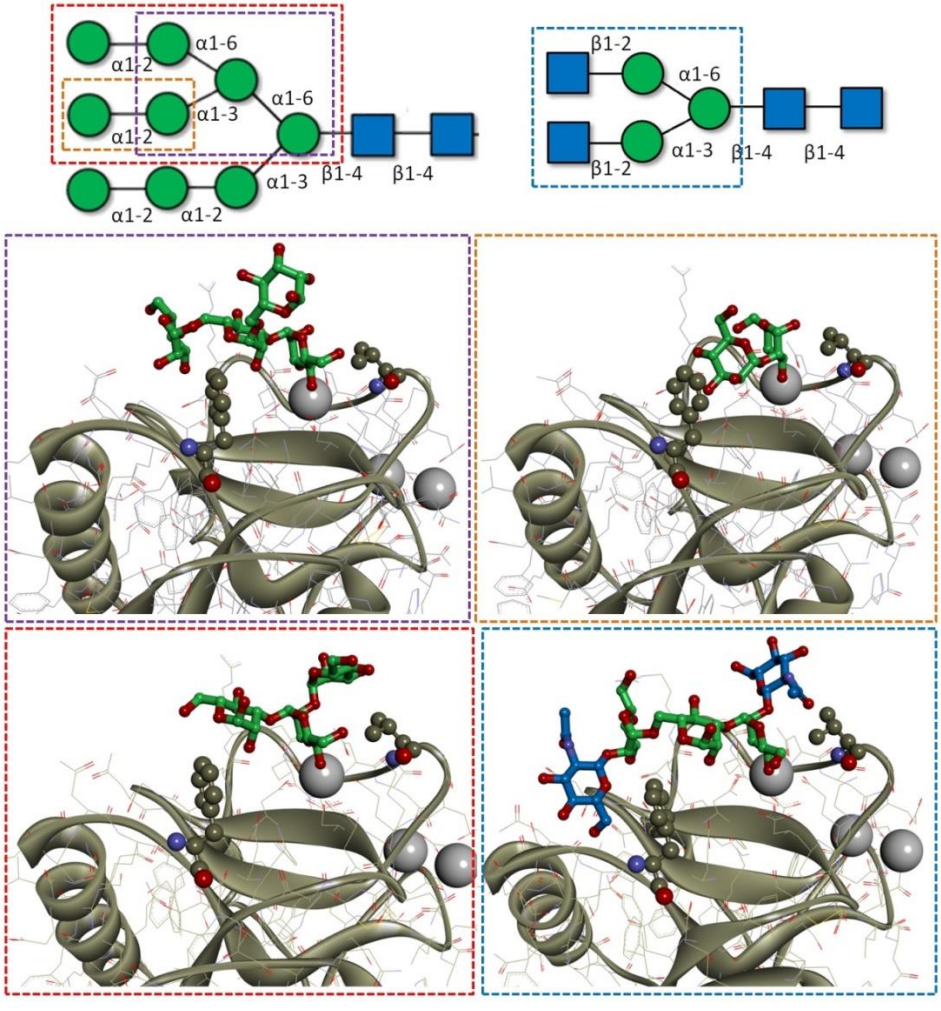


Figure 1.17. *The four available X-Ray models describing the interaction of DC-SIGN with different oligomannose fragments, namely,* Man_4 *(PDB: 1SL4, purple box),* Man_2 *(PDB: 2IT6, orange box)* Man_6 *(PDB: 2IT5, red box)* $GlcNAc_2Man_3$ *(PDB: 1K9I, blue box). Only the major binding poses are represented. The sidechains of Phe313 and Val351 are detailed.*

In this scenario, X-Ray models as well as NMR analyses have provided extensive data about the molecular basis underlying this fine selectivity [207-208]. Crystal structures containing the Man_3 core always display the α1-3-linked residue directly coordinated to the calcium ion. The central and the outer α1-6-linked mannoses lie along the polar surface defined by β_3 and β_4 strands, flanking the aromatic ring of Phe313 without originating steric clashes (Figure 1.17). In the X-Ray model with $GlcNAc_2Man_3$, one GlcNAc residue is located onto the long loop, presumably establishing favorable hydrophobic contacts with the methyl groups of Val351, whereas the other one fits into a hydrophilic pocket defined by the β-sheet turn and the α_2-helix [207]. Interestingly, the further elongation of these glycans, for instance with Gal or sialic acid, negatively affects the recognition of the core Man_3 [209] (Figure 1.18). Moreover, the binding ability seems to be partially recovered when the GlcNAc moiety is bearing GalNAc rather than Gal [210] (Figure 1.18, *entries 12 and 15*). Unfortunately, there is not a structural basis to explain that yet.

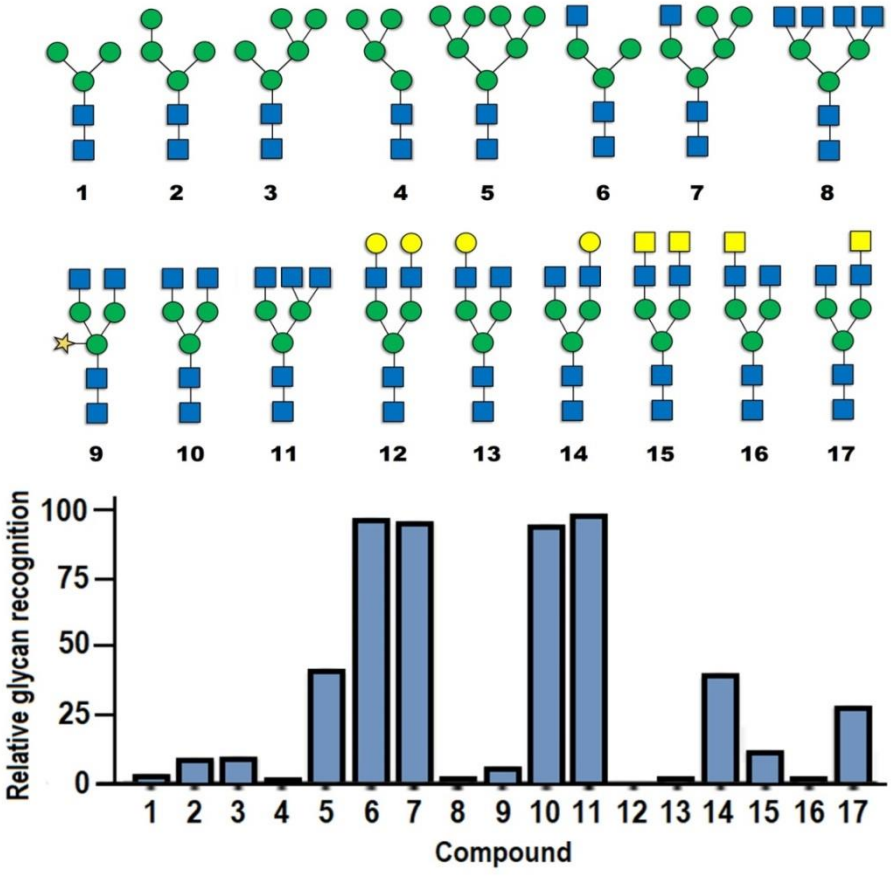


Figure 1.18. *Relative affinities determined by glycan arrays for 17 Man-containing oligosaccharides. Adapted from [211].*

Notably, the CRD-Man_4 complex demonstrates that the central unit of Man_3 should be always α-linked to the upstream scaffold to make the binding possible through this trimannoside core. The β-anomer would place the linked sugar too close to Phe313, whereby residue accommodation would not be viable (Figure 1.17). This presumably explains why DC-SIGN is completely unable to bind to the $Man_3GlcNAc_2$ fragment, and also underlines the double role of Phe313 as a structural element for surface complementation and for discrimination between the inner and the outer Man_3 cores present in complex oligomannans [196]. Very interestingly, recent array studies with a large library of Man_3-containing complex N-glycans have raised new issues about the cooperative influence of all the surrounding residues in the binding to this trimannose core [211]. As highlights, it has been shown that the binding of β1-4-linked Man_3 is actually possible whether the $Man_3GlcNAc_2$ pentasaccharide is elongated with two additional GlcNAc units (Figure 1.18, *entries 1 and 10*), and also, further asymmetric galactosylation seems to be solely tolerated at the α1-6 arm (Figure 1.18, *entries 12-14*).

The Manα1-2Man fragment can be efficiently inserted in many ways onto the binding site region, opening the door to multiple plausible binding poses. Both X-Ray and NMR data have pointed out the preferential coordination of the inner mannose, while experimental STDs (in combination with CORCEMA-ST) using the simple disaccharide have actually proven that there must be an additional minor binding mode co-existing at the same time, in which the Ca^{2+} ion is targeted by the outer Man [212]. Fittingly, crystallographic models with the same fragment support that the non-reducing Man can perfectly pack against the aromatic ring of Phe313, thus justifying the NMR data (Figure 1.17). Alternatively, the crystal containing the Man_6 fragment also reveals that the Man can adopt the opposite presentation, with the ring rotated 180º. In this case, the terminal Man points towards Val351, and the core is placed similar to those of Man_4 and $GlcNAc_2Man_3$ fragments [213]. All these binding modes are compatible with the Man_9 architecture, as assessed by docking models.

In regard to the mannan affinities, the low μM dissociation constants observed for $Man_9GlcNAc_2$ and other highly-mannosylated N-glycans are now reasonably explained whether considering the expected *statistical rebinding* effects: the presence of several Man residues that, in some cases, could display two ring orientations. Accordingly, these effects justify that mannan fragments usually exhibit larger binding constants as their size increases (number of Man residues) [196,214-215].

However, even considering the favorable statistical re-association, more complex spatial factors should be taken into account in the recognition of high-Man glycans, especially those regarding sugar accessibility and sugar density. These factors strongly depend on the glycan flexibility and conformational space, and may explain some discrepancies found when comparing affinities between oligomannans [215-216], namely changes in the global affinity that do not correlate with the expected effect derived from the addition or removal of available Man residues. Three-dimensional effects affecting glycan recognition become more evident in biological studies, when considering the whole carbohydrate-containing entity, either the isolated pathogen glycostructures or the entire organism in "*in vitro*" studies. As an example, dissimilar interactions between DC-SIGN and *Mycobacterium* species have been ascribed to the variable number and distribution of Man caps on the surface-exposed ManLAMs [217]. Alike, it is well known that *Candida* species target DC-SIGN through their N-linked mannans, whereas O-linked mannans and phosphomannans barely participate in the pathogen uptake [112]. Despite their irrelevant role in binding, phosphomannans are thought to influence the conformational space of the N-linked mannans, as depletion of the former affects the binding strength of the latter [218]. Unfortunately, addressing a detailed structural analysis from such a high level of complexity is unrealistic and unworkable for now. Also, each glycan scaffold adopts a suitable binding mode and may trigger local conformational changes in the receptor that are transmitted to distal regions, what will eventually have biological implications like influencing intracellular lectin signaling. There are examples in the literature related to DC-SIGN with gp120 [219] and L-SIGN with $GlcNAc_2Man_3$/$Man_9GlcNAc$ [220].

1.4.2. Fucose-based binding

DC-SIGN binding to Fuc also drew attention of researchers since the biological importance of this receptor was uncovered. This sugar naturally appears in a wide range of self- and non self-antigens with different scaffolds, mainly in **Lewis antigens and Blood group antigens** (See *Chapter 4*). In comparison with mannans, these fucosylated motifs display a greater structural variability regarding the surrounding residues (usually Gal, Glc, their N-Acetyl derivatives, sialic acid and other fucoses) and the type of glycosidic linkages existing between them. Over the years, research data have evidenced highly variable affinities for fucosylated antigens. Often, due to the lack of enough structural data, these results have exposed unexplained discrepancies and vague correlations that could have not been accurately ascribed to structural

factors. Overall, the recognition of Fuc-containing glycans perfectly reflects the strong influence of the ligand presentation in the binding process.

Mostly from array data [196,214-215,221], DC-SIGN CRD is an acknowledged receptor of Lewis-type antigens (Le^X, Le^Y, Le^A , Le^B, LDN-related scaffolds) in which the fucose is α1-3 or α1-4-linked to a central GlcNAc residue (Figure 1.19). It also interacts with blood group antigens (BgA, BgB, BgH) with a core galactose bearing an α1-2-Fuc moiety. By contrast, the negative effects that sialylation and sulfation frequently exert in modified Lewis-type scaffolds are well documented: the elongation of the non-reducing end with a sialic acid unit almost completely abrogates the binding in most of cases, whereas the presence of sulfate groups can entail either negligible or negative consequences for binding, depending on the sulfated position and the type of fucosylated scaffold (Figure 1.19, *entries 7-8 and 13-14*).

From now on, these general guidelines occasionally fail to explain some other differences found between related structures, especially when considering elongated fragments with more neighboring sugars. Some noticed patterns are:

- Sequential repeats of a Lewis antigen (e.g. di-Le^X, tri-Le^X, di-LDNF) expectedly lead to increased affinities, likely contributed by additional "*statistical rebinding*" effects. However, the depletion of individual fucose moieties in these oligomers has an unclear effect: in Le^X-containing structures, the removal of the terminal Fuc sharply decreases the affinity, suggesting a minor role in the lectin recognition from the fucoses located at inner positions. On the contrary, the same affinity drop is observed in Le^A repeats when an inner Fuc residue is cleaved while maintaining the outer one (Figure 1.19, *entries 4-6 and 10-12*).
- In Lewis-type antigens, the substitution of the non-reducing Gal moiety by GalNAc is nearly irrelevant for binding, but it has a noticeable effect when it comes to the interaction with the blood groups, as the B antigen (BgB, contains Gal) exhibits a better affinity than the A antigen (BgA, contains GalNAc) (See *Chapter 4*).
- The complete absence of the terminal Gal in blood group antigens (as in the H antigen, BgH) gives rise to a poor binding, suggesting that the role of this non-reducing sugar may be more relevant for the blood groups than for the Lewis groups. In fact, structural modifications affecting to this residue, such as changing the anomer (α-Gal/β-Gal) or the type of

glycosidic linkage (α1-3/α1-4) have a negative impact on the interaction, as a general rule (Figure 1.19, *entries 16-21*).

- The type of glycosidic bond that attaches the central sugar to the glycan scaffold may condition the sugar accommodation at the binding site. Roughly, the most preferable linkage for both Lewis and BG antigens is β1-4, as those glycans bearing β1-3-linked ones tend to show lower affinities.

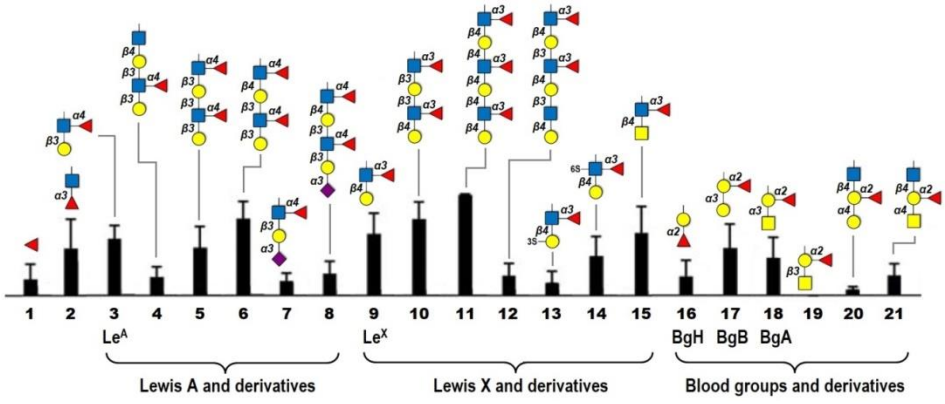


Figure 1.19. *Relative affinities determined by glycan arrays for 21 related Fuc-containing oligosaccharides. Adapted from [215].*

Obviously, all these queries are meant to have a structurally supported reasoning, but there are still very few pieces of information about the molecular recognition of Fuc-containing oligosaccharides by DC-SIGN. Essentially, X-Ray and NMR techniques have been applied to unveil the molecular basis for the recognition of the Le^X antigen [196,222]. Both sources of data agree that the Fuc residue is coordinated to the metal ion through its OH3 and OH4 groups, while the Gal ring is placed into the flat area defined between Phe313 and the Ca^{2+}, establishing a secondary H-bond interaction with Asp367 through OH6 (Figure 1.20). Additionally, a thorough analysis of the sugar positioning has allowed hypothesizing about the recognition of related structures:

- The presence of GalNAc instead of Gal does not presumably produce severe changes in the binding, as the N-Acetyl group at carbon 2 would be expected to point away from the protein surface. Also, the attachment of an additional Fuc moiety to the same position could be likewise tolerated. Taken together, these observations support the binding of LDNF and Le^Y motifs to DC-SIGN, respectively.

- In the same line, the recognition of sialyl-Le^X (sLe^X) would be expected to fail due to foreseeable clashes between the sialic acid moiety and Phe313. Strikingly, the mutant F313A neither recognizes sLe^X, supporting that the stabilization of the sialic acid ring in selectins [223], for instance, requires the existence of additional specific electrostatic and vdW interactions arising from the protein environment. These interactions might not be present in DC-SIGN.

On the other hand, the NMR results have highlighted the crucial role of the hydrophobic interactions involving Val351 in the stabilization of DC-SIGN-fucose complexes. The H2 of Fuc establishes a tight van der Waals contact with the Val sidechain, as assessed by STD experiments. In parallel, mutagenesis studies have underlined the role of this aliphatic residue in Fuc selectivity [195-196]: replacing Val351 by a glycine residue abrogates the recognition of fucosylated ligands, whereas Man can still bind. Instead, the swapping of Val by Ser turns DC-SIGN selectivity into that of L-SIGN (Ser363), and Le^X can be no longer bound. Besides Fuc protons, the N-Acetyl group of the core GlcNAc lies close to Val351 and the experimental STD values have confirmed that it notably contributes to the hydrophobic stabilization as well.

Moreover, the models built using experimental NMR constrains (STD and ligand transferred NOEs) have shown discrepancies with the crystal structure. Specifically, STDs agree with closer contacts between the Gal ring and the protein surface while the N-Acetyl group of GlcNAc should be located fairly close to Val351. Both observations point out the importance of these secondary interactions for the binding stability. In addition, the bound conformation of the Lewis-type trisaccharide interestingly seems to be slightly more “opened”, different from that in solution. This is partly surprising given the known conformational rigidity of histo blood groups [224-226] even when bound to their receptors [9,184,227].

Other fucosylated oligosaccharides, like Le^A or pseudo-Le^Y (the second Fuc is attached to OH3 of Gal), have been investigated to a lesser extent, only making use of docking models. In the Le^A-DC-SIGN complex [195], the central GlcNAc is rotated 180º because of the different configuration of glycosidic linkages. Thus, Val351 would be now contacting the CH_2OH group rather than the N-Acetyl group, and the hydrophobic interaction would be maintained yet, although expected to be weaker due to the presence of the polar OH6 end (Figure 1.20). Interestingly, this polar group would contact Ser363 in L-SIGN,

explaining why this C-type lectin still recognizes Le^A and Le^B. Nevertheless, this arrangement does not support Le^Y recognition by L-SIGN.

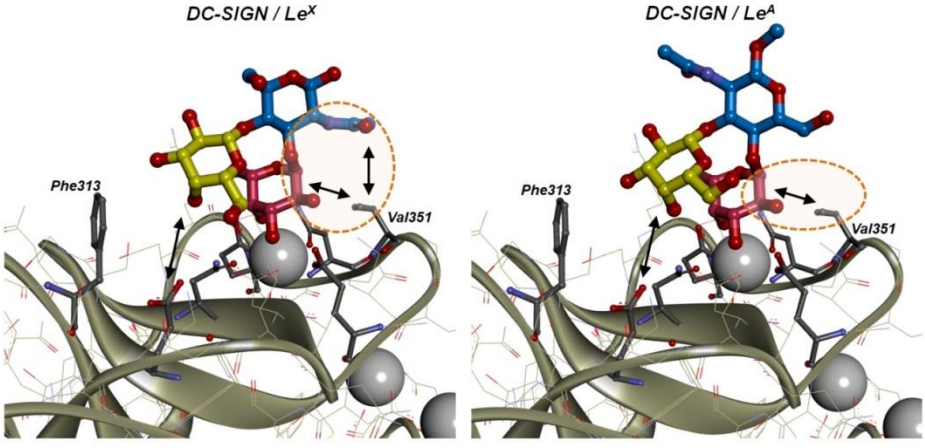


Figure 1.20. *On the left, the DC-SIGN-LNFP III complex (1SL5) displaying the accommodation of the Lewis X antigen at the primary binding site. On the right, a simulated model of the interaction of DC-SIGN with the Lewis A trisaccharide. The Le structure has been modeled with the GLYCAM server [228] and manually docked into the calcium site by superimposing both interacting Fuc moieties. As noted, the GlcNAc residue is rotated 180º in the latter case. Color code: Fuc: red. Gal: yellow. GlcNAc: blue.*

The interaction of DC-SIGN with pseudo-Le^Y is as strong as with Le^Y [229]. Therefore, the protein-ligand complex is thought to be similar to that of Le^X and Le^Y, although the docked models suggest that the attachment of the second Fuc to Gal C3 instead of Gal C2 would force the aromatic ring of Phe313 to adopt a different conformation, avoiding possible steric clashes.

1.4.3. Artificial receptors: towards the development of therapeutics

DC-SIGN has become a very promising target for the development of therapeutic methodologies against diseases in which this receptor takes part. Aside from anti-glycan vaccines or cell immunotherapy, one of the most employed strategies in medicine consists on blocking the receptors directly involved in the pathologic disorder [230]. Indeed, nature itself has provided mammals with a plethora of oligosaccharides in the human milk to block or modulate C-type lectin-driven interactions and protect the breastfed infants from external microorganisms [231]. Actually, human milk contains glycoproteins with the Le^X motif which are able to block DC-SIGN, hence repelling the pathogen attachment [232-233].

There are few recent examples about glycan-containing biostructures, like immunomodulators, nanoparticles or liposomes, which have successfully worked as anti-glycan vaccines to simultaneously target DC-SIGN and stimulate DC activation [234-236]. Apart from those, the rational design of glycomimetics keeps being the major methodology to address the quest for better substrates than the natural ones. In spite of the aforementioned structural limitations that are common for C-type lectins, good inhibitors have been developed along this decade, surpassing the affinity of natural monosaccharides above 3 orders of magnitude. Then, these improved mimetics can be arranged into a multivalent presentation, eventually yielding incredibly potent inhibitors [237].

- **Exploiting secondary interactions**

The design of Man-based mimetics has been essentially oriented to exploit additional van der Waals contacts around the binding site, using the Man scaffold to drive the coordination to the calcium ion and bear other groups or modified sugars. The earliest research already proved that a simple structural modification at position 2, like introducing a methylamino group, can increase the affinity for Man more than one order of magnitude (K_i = 0.35 mM) [238]. Mostly, the use of carba-analogues (cyclohexane instead of pyranose) has accompanied this research drift in order to enhance the drug-like character of designed mimetics (increased lipophilicity) and their metabolic stability against enzymatic degradation [239-240].

In this line, a great deal of work has focused on improving the structural features of the Manα1-2Man disaccharide (Figure 1.21A). The substitution of the reducing Man by a functionalized cyclohexane ring provided one of the first and most studied mimetic scaffolds [241]. Interestingly, X-Ray and NMR studies confirmed that it adopts a completely new binding pose where the inner cyclohexane is stacked on top of Val351 rather than lying near to Phe313, as seen for the natural disaccharide [242] (Figure 1.21B). Moreover, this alternative hydrophobic stabilization would explain the improved selectivity for DC-SIGN as compared to Langerin (4-fold better). In this scenario, the subsequent substitution of the ester moieties by arylamide groups reinforced the binding with additional contacts on the back part of the long loop [243-244], providing higher affinities in the μM range. Additionally, these modifications hampered the recognition by Langerin, increasing the selectivity towards DC-SIGN even more. Recently, a total selectivity for DC-SIGN has been achieved by combining these arylamide groups and positively

charged groups in the same molecule [245]. The presence of a positive substituent in the interacting Man ring precludes the binding to Langerin as a result of repulsive electrostatic contacts with nearby lysine sidechains. At the same time, this positive functional group has been subsequently modified to also provide stabilizing contacts with the neighboring sidechains of DC-SIGN (near to Phe313), yielding more potent mimetics [246].

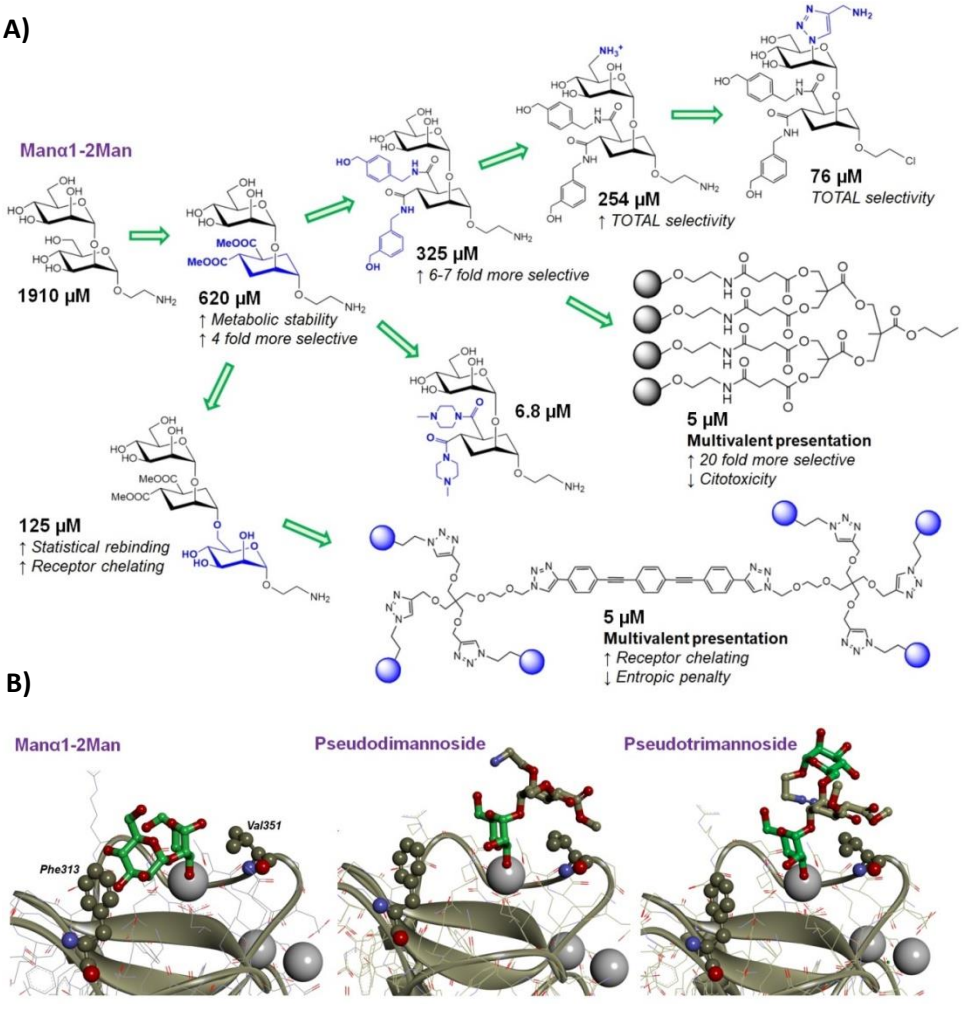


Figure 1.21. *A) Structural modifications progressively performed on the Manα1-2Man scaffold over the last decade to improve its inhibitory potency and enhance its selectivity towards DC-SIGN, minimizing its recognition by Langerin. B) Crystallographic models of DC-SIGN in complex with the pseudodimannoside and the pseudotrimannoside mimetics (PDBs: 2XR5 and 2XR6, respectively), as compared with the major binding pose reported for the Manα1-2Man fragment. As can be observed, the interacting Man adopts in all cases the same ring orientation.*

A similar line of research tried to target the opposite secondary pocket with an elongated mimetic bearing a third natural Man residue at the reducing end [247]. The resulting compound was one order of magnitude more potent than the previous dimannoside mimetic, but surprisingly, the major binding mode was again identical and the affinity upgrade simply came from the existence of a minor additional binding pose involving the inner Man residue [248] (Figure 1.21B). Indeed, at low ligand/protein ratios, this mimetic has been noticed to target two independent binding sites, bridging two separated DC-SIGN tetramers in solution [249].

The design of multivalent systems containing these improved analogues has developed in parallel. A tetravalent presentation enables additively combining the individual affinities of the monovalent mimetics with multivalence effects [250]. Furthermore, these dendrons were readily affordable and displayed better biological properties, including high inhibitory potencies against HIV, selectivity for DC-SIGN and low cytotoxicity [251]. However, the global affinities become less predictable in higher order glycoconjugates, since additional effects like sugar density/accessibility or CRD chelation begin to take effect [252]. Specifically, the individual contribution of each epitope is diluted, and similar affinities have been reported for 32-mer dendrons containing the dimannoside and the trimannoside mimetics, respectively [253]. A different design with a rigid rod-shaped core and flexible spacers has enabled exploring those multivalence factors in a controlled fashion, using a hexavalent presentation [254-255]. Longer rod-like cores contributed to diminish the entropic penalty, while the spacers of variable length allowed modulating the epitope accessibility and the ability to chelate several CRDs.

Other works have directly attempted to create stabilizing hydrophobic contacts with Phe313 dispensing with the neighboring sugar residues [256]. Using a glycerol chain as a spacer, Kotar et al [257] have attached aromatic moieties to the anomeric carbon to develop a series of mimetics with affinities in the low micromolar range. Experimental STDs, assisted by docking models, have assessed the good interaction provided by a naphtyl moiety contacting the aromatic ring of Phe313 (Figure 1.22), while another aromatic entity (napthyl or benzyl) is meant to be placed into a different pocket close to Asn344/Arg345.

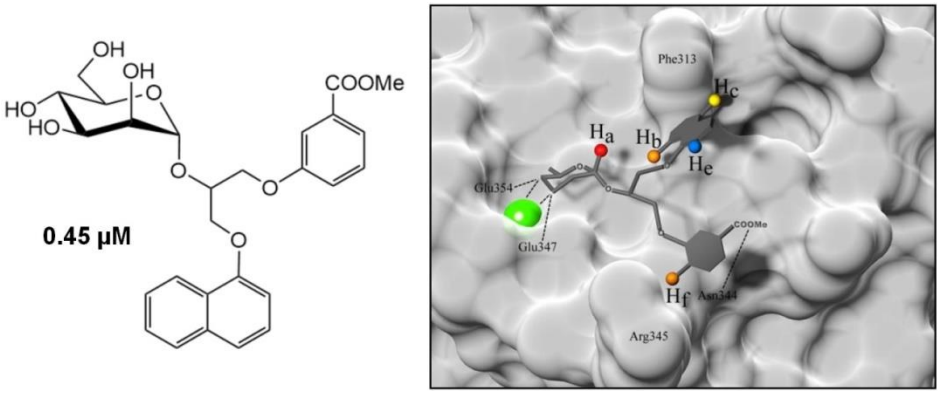


Figure 1.22. *Most potent inhibitor obtained through the introduction of non-sugar substituents at the anomeric position of the recognized Man. On the right, modeling based on measured STD effects. The Ca^{2+} ion is shown as a green sphere. Some ligand protons whose STDs have been measured are indicated and colored based on ligand-receptor inter-proton distances (from red to blue). Adapted from [257].*

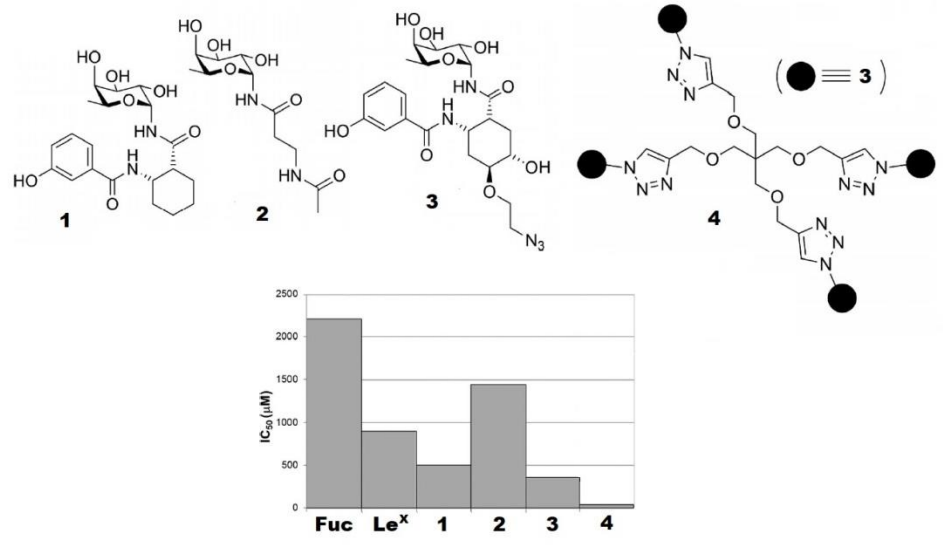


Figure 1.23. *Some Le^X analogues and their relative inhibitory potencies as compared with those of natural Fuc and Le^X antigens. The best designs are those of **1** and **3**, reflecting the importance of the aromatic moiety mimicking the Gal unit (better than **2**). Adapted from [260].*

Regarding Fuc mimicking, few works have described the synthesis of Lewis X analogues with similar inhibitory potencies, in which both GlcNAc and Gal residues are substituted by cyclohexane rings [258]. According to crystallographic data, the Gal mimetic seems quite important for assisting Fuc binding, whereas modifications on the central cyclohexane barely affect the

affinity positively [259]. Remarkably, the best mimetic has been achieved after substituting the cyclohexane stereochemistry of the Gal residue by a planar aromatic ring [260] (Figure 1.23). NMR data (STD) have confirmed the involvement of this phenyl group in close contacts with the protein and docking analyses have strongly suggested that this moiety actually contacts the sidechain of Phe313 [261].

- **Exploiting multivalence**

As mentioned at the beginning of this section, multivalent effects are crucial for enhancing the poor affinities of monosaccharides or small oligosaccharides for DC-SIGN, and for any C-type lectin in general [191,262]. Notably, this is a field of intense research since the factors that govern multivalent effects are not so straightforward, and the design of potent inhibitors require a thorough analysis of geometrical and thermodynamic aspects including the type of core or skeleton, the chemical nature of the spacers, as well as their rigidity or flexibility; the closeness between interacting epitopes (sugar density and accessibility), and the nature of the proper glycan moieties. This issue is even more complex whether considering that binding potencies are not necessarily correlated with internalizing capabilities: the best multivalent ligand may not be suitable for DC activation after being recognized [263].

At last, the multivalency enhancement of a glycoconjugate is given by the so-called "*affinity per residue*": how higher the affinity of an individual epitope in a multivalent scaffold is as compared with the single epitope itself.

The first Man-containing glycoconjugates already proved to be effective for avoiding pathogen attachment in "*in vitro*" studies, efficiently competing with the natural viral glycoproteins (IC_{50} = 0.34 µM for EBOV; 50 µM for HIV) [264-265]. In principle, this affinity improvement came exclusively from chelating effects when interacting with DC-SIGN tetramers, as the multivalent effects disappeared in the presence of the single CRD [266] (Figure 1.24). The distribution of the sugar epitopes throughout the glycoconjugate surface also affects the global potencies, and many designs have reported such effects [267]. Polyamide scaffolds bearing alkyl oligomannans (alkyl-Man_4, alkyl-Man_9) have provided increasing affinities according to the rising number of mannoses from the first (3 mannoses) to the second generation (9 mannoses). However, there was not a significant change from the second to the third generation (27 mannoses), whereby the second generation was good enough to imitate the glycan density of gp120 on the surface of the array plate [268]. Similarly, the best affinities reported for golden core glyconanoparticles

coated with mannosides have been achieved for a 50% of ligand occupation [269-270]. Simultaneously, these systems also illustrate the relevance of the monovalent epitope in regard to its accessibility. Thus, although Man_5 is expected to better bind DC-SIGN than the linear Manα1-2Man disaccharide, the last one provides higher potencies in a multivalent presentation, likely because it preserves all its binding modes whereas Man_5 is spatially more constrained. A potential solution to this problem consists of using longer spacers to increase the ligand dispersion, as has been suitably applied for the preparation of PEG-based polymannosides with very high affinities per Man residue (12000-fold) [271] and glycofullerenes with subnanomolar affinities [272-273] (Figure 1.24).

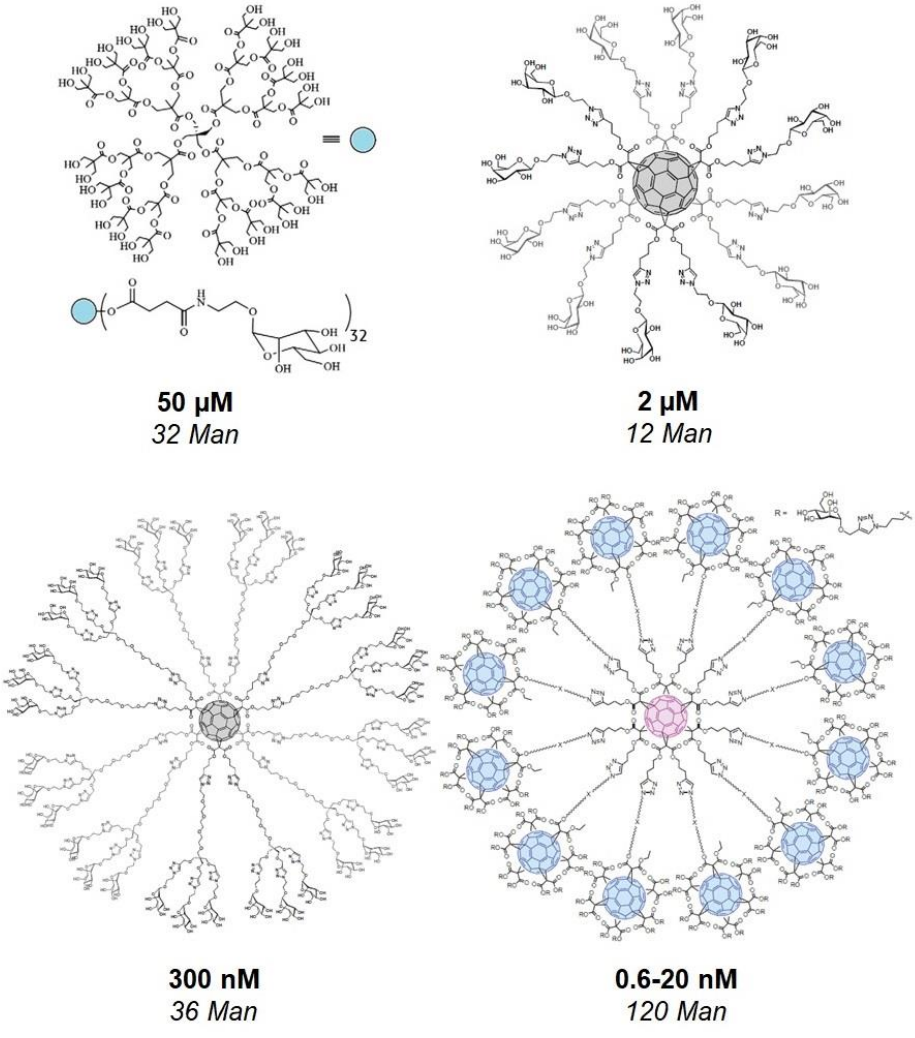


Figure 1.24. *Most potent glycodendrimers developed to target DC-SIGN. Adapted from [266,272-273].*

The presentation of the ligand strongly modulates the affinities as well. For instance, the alternative coupling of sugar monosaccharides through its C6 position instead of the anomeric carbon seems to be irrelevant when building large glycoconjugates [274]. On the contrary, the substitution of the O-linkage by a C-linkage can exert either negative or positive effects, depending on the specific case. Roughly, C-fucosides have been reported to be beneficial for binding [275], whereas C-mannosides usually produce less relevant or even negative changes in the affinities [276]. In certain cases, sugar clusters located at a focalized point of a rigid glycoconjugate help to enhance the binding, increasing the probability of statistical rebinding phenomena and easing the targeting of two or more neighboring CRDs at once [277]. As an example, glycolipids with long enough aliphatic chains can trigger the formation of micelles by self-association and display multivalent behaviors [278] (Figure 1.25A).

Globular neoglycoproteins and linear glycopeptides constitute another source of multivalent architectures extensively exploited for many years [32,279]. In fact, these tools have remarkably contributed to the recent development of very potent inhibitors (nM-pM) to prevent the DC-SIGN-mediated binding of the envelope GP of EBOV and gp120 of HIV [280-281]. In the last case, DCs can internalize gp120 yet due to the participation of the co-receptor CD4. Nevertheless, these results are promising for expanding and optimizing their applications in the biological scope. An important advantage of linear glycopeptides is that the amount of residues can be increased without affecting the sugar density, just elongating the peptidic chain. The same benefit has been also exploited by employing DNA/RNA templates instead of peptidic chains [282-283]. Thus, a linear dependence between the degree of sugar coating and the affinity is readily achievable, and the best potencies have been reported in most of cases for 100% of epitope occupancy [284-285] (Figure 1.25B). Moreover, the linear geometry enables the creation of chelation effects more easily. Interestingly, in glycopolymers with intermediate percentages of glycan substitution, the distribution pattern may affect the affinity [286].

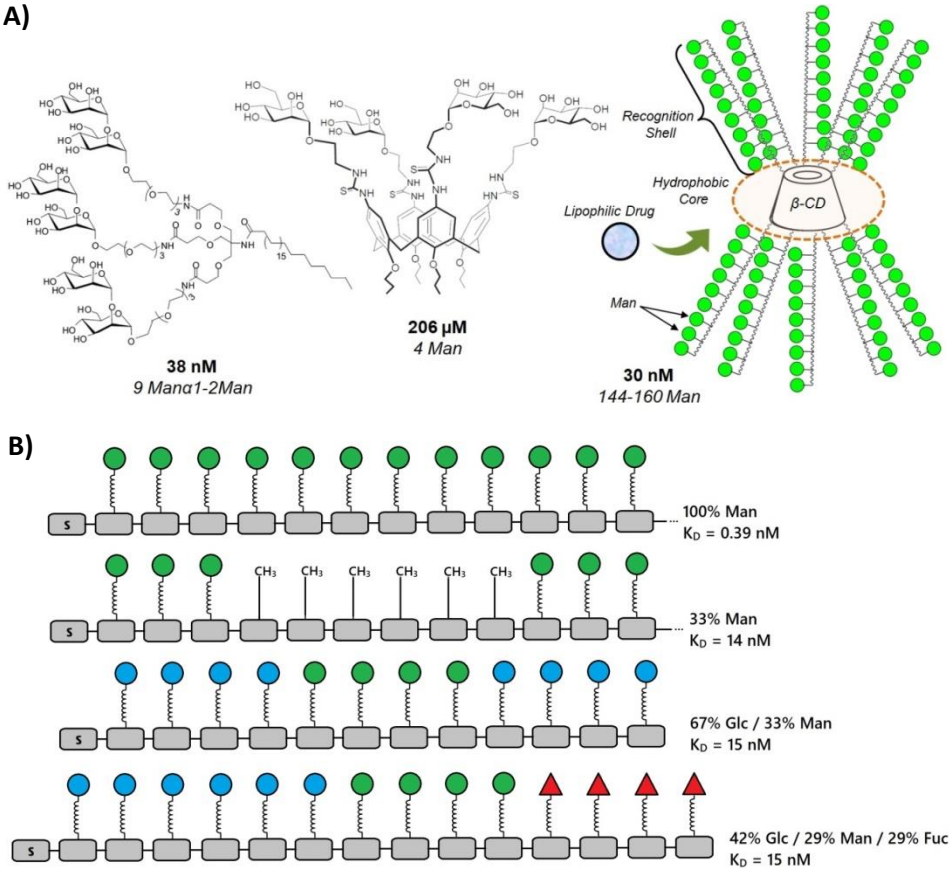


Figure 1.25. *A) Most potent glycoclusters developed to target DC-SIGN. From left to right, the shown structures have as base scaffold a lipid chain [278], a calyx[4]arene [277] and a 6–cyclodextrin [287], respectively. B) Linear glycopeptides with different epitope distributions designed to disrupt the DC-SIGN-gp120 interaction [286]. The "S" block corresponds to the starter structure (no epitope). Each block bears a unique monosaccharide. The first two designs on top are made of 23 and 27 blocks, respectively (only a fragment shown).*

Finally, it is worth mentioning liposomes as multivalent scaffolds carrying multiple sugar copies. Liposomes are commonly used as carriers for intracellular drug delivery and, in this regard, glycoliposomes have been also employed for this purpose aimed at targeting DC-SIGN, among other lectins [288-289]. Strikingly, the ligand presentation plays here an unclear role, since both DC-SIGN and Langerin recognize Le^{Y}-coated liposomes, but whereas DCs can internalize them via DC-SIGN, Langerhans cells cannot do so via Langerin. Instead, glycopolymers are actually effective for drug delivery in Langerhans cells [290]. This work well reflects how subtle variations in the glycan presentation and accessibility can trigger deep differences in the recognition by related lectins. In this line, cyclodextrins decorated with polymannose

branches have been under study as well [287] (Figure 1.25A). However, although the attached branches improve the potencies of these cyclodextrins below the micromolar range, the carrier capability remain limited due to the additional difficulties to load the tested drugs into the small, hardly accessible hydrophobic core.

- **Other non-sugar inhibitors**

Finally, in spite of the low druggability featuring most of C-type lectins, several research groups have efficiently found out non-carbohydrate inhibitors by ligand screening (Figure 1.26). Some of them, like those based on the shikimic acid scaffold, exhibit affinities very similar to these of the natural monosaccharides (Figure 1.26, *right*). Further structure improvements have yielded better inhibitors which display low micromolar potencies (about 3 µM) when loaded onto a heptavalent polymer [291]. NMR studies have assessed that these mimetics specifically interact at the primary Ca^{2+} site, and the conjugation of multiple copies to a bovine serum albumin (BSA) scaffold allow them to promote cell signaling by DC-SIGN targeting [292]. Another set of interesting non-natural inhibitors have been discovered by solid phase arrays. Curiously, none of the described inhibitors had polar groups with the right stereochemistry to chelate the calcium ion, but even so several of them reached low micromolar affinities [293] (Figure 1.26, *left*). Structurally, various candidates shared a common core composed of two fused rings (at least one aromatic and/or one heterocycle). The optimization of this core with other suitable aromatic functional groups at remote positions gave rise to very good inhibitors with submicromolar potencies [294] (Figure 1.26, *top*).

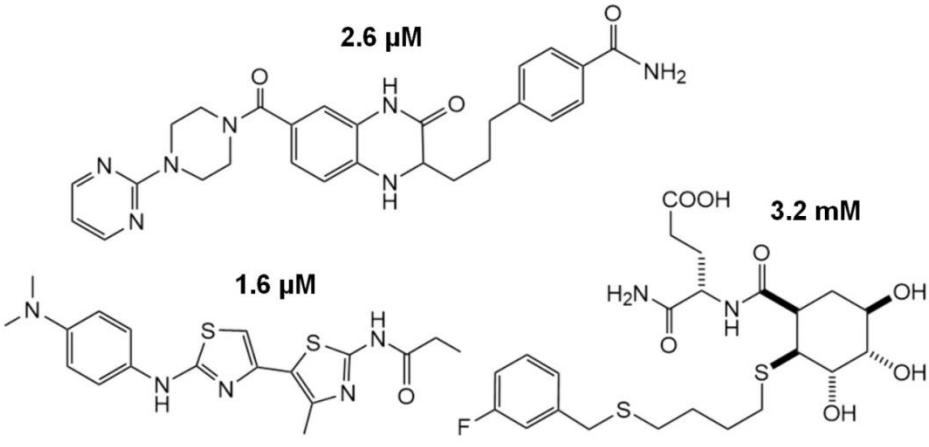
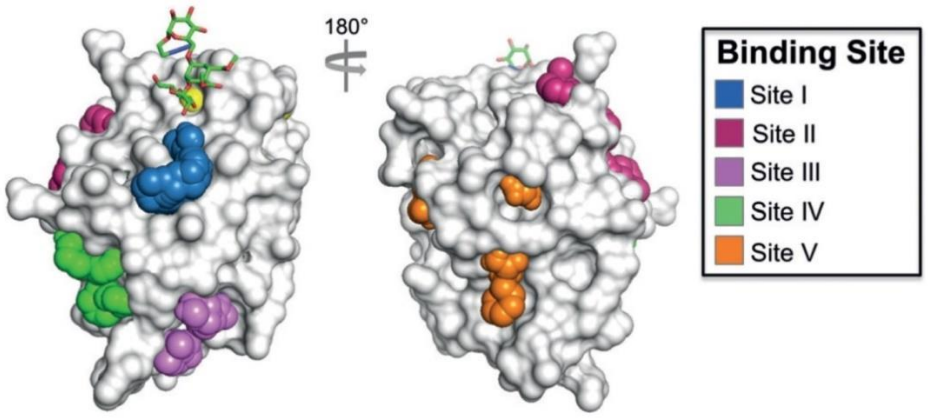


Figure 1.26. *Most relevant non-sugar inhibitors discovered for DC-SIGN.*



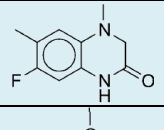
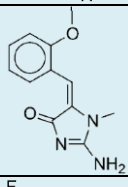
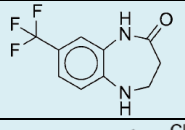
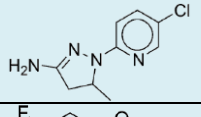
	Screen	K_d (mM)[a]	K_d (mM)[b]	Site
	NMR	0.3 ± 0.2	0.5 ± 0.1	**I**
	SPR	-	0.4 ± 0.1	**II**
	NMR, SPR, array	1.0 ± 0.1	1.8 ± 0.3	**III**
	SPR	-	0.7 ± 0.1	**IV**
	NMR, array	0.4 ± 0.1	0.7 ± 0.1	**V**

Figure 1.27. *The five potential binding sites that can accommodate drug-like molecules in the DC-SIGN CRD, as indicated by computational docking. Below, five active fragments found by screening using NMR, SPR and/or arrays, and the site of interaction expected for each one as predicted by the docked models. K_ds have been determined in the presence of EDTA (a) or Ca^{2+} (b). Adapted from [298].*

More recently, fragment screening methods combining SPR, ^{19}F-NMR and other techniques like reporter displacement assay (RDA) have been applied to the quest for new drug candidates [295-297]. In parallel, NMR-based titration methods have evidenced that many of the active fragments do not interact at

the primary binding site. Instead, computational methods have suggested the existence of five different druggable sites distributed along DC-SIGN surface, where these compounds would presumably bind [298] (Figure 1.27). Thus, these findings have shed light into the design of new allosteric, non-sugar inhibitors to potentially control the glycan attachment to this lectin.

1.5. References

1. G. MacPherson and J. Austyn. The Immune System. In *Exploring Immunology: Concepts and Evidence*, eds. G. MacPherson and J. Austyn. Wiley-VCH Verlag GmbH & Co. KGaA (Weinheim), 1st edn, 2012, **ch. 1**.
2. F. Bacon. Introduction to the Immune System. In *Primer to the Immune Response*, eds. T. Mak, M. Saunders and B. Jett. Academic Cell (Boston), 2nd edn, 2014, **ch. 1**.
3. C. A. Janeway Jr. and R. Medzhitov. Innate Immune Recognition. *Annu. Rev. Immunol.,* (2002), **20**, 197–216.
4. R. M. Steinman, in *Fundamental Immunology*, ed. W. E. Paul. Lippincott Williams & Wilkins (Lippincott-Raven) (Philadelphia). 7th edn, 2017.
5. S. McComb, A. Thiriot , L. Krishnan and F. Stark. Introduction to the Immune System. In *Immunoproteomics: Methods and Protocols*, eds. K.M. Fulton and S. M. Twine. Springer Science+Business Media (Luxembourg), 2013, **ch.1**. *Methods Mol. Biol.,* (2013), **1061**, 1-20.
6. Y. J. Liu. Dendritic cell subsets and lineages, and their functions in innate and adaptive immunity. *Cell,* (2001), **106**, 259-262.
7. Y.-J. Liu, H. Kanzler, V. Soumelis and M. Gilliet. Dendritic cell lineage, plasticity and cross-regulation. *Nat. Immunol.,* (2001), **2**, 585–589.
8. A. K. Abbas, A. H. Lichtman and S. Pillai. Cells and Tissues of the Immune System. In *Cellular and Molecular Immunology*, eds. A. K. Abbas, A. H. Lichtman and S. Pillai. Elsevier (Amsterdam), 9th edn, 2017, **ch.2**.
9. J. Banchereau and R. M. Steinman. Dendritic cells and the control of immunity. *Nature,* (1998), **392**, 245–252.
10. E. Ingulli, A. Mondino, A. Khoruts and M. K. Jenkins. *In vivo* detection of dendritic cell antigen presentation to CD4+ T cells. *J. Exp. Med.*, (1997), **185,** 2133–2141.
11. S. A. Luther, A. Gulbranson-Judge, H. Acha-Orbea and I. C. M. Maclennan. Viral superantigen drives extrafollicular and follicular B differentiation leading to virus-specific antibody production. *J. Exp. Med.,* (1997), **185**, 551–562.
12. C. Caux, B. Vanbervliet, C. Massacrier, C. Dezutter-Dambuyant, B. de Saint-Vis, C. Jacquet, K. Yoneda, S. Imamura, D. Schmitt and J. Banchereau. CD34+ hematopoietic progenitors from human cord blood differentiate along two independent dendritic cell pathways in response to GM-CSF+ TNFα. *J. Exp. Med.*, (1996), **184**, 695–706.

13. a) D. Strunk, C. Egger, G. Leitner, D. Hanau and G. Stingl. A skin homing molecule defines the Langerhans cells progenitor in human peripheral blood. *J. Exp. Med.*, (1997) **185**, 1131–1136. b) T. Ito, M. Inaba, K. Inaba, J. Toki, S. Sogo, T. Iguchi, Y. Adachi, K. Yamaguchi, R. Amakawa, J. Valladeau, S. Saeland, S. Furuhara and S. Ikehara. A CD1a+/CD11c+ subset of human blood dendritic cells is a direct precursor of Langerhans cells**.** *J. Immunol.*, (1999)**, 163**, 1409–1419.
14. B. Dubois, B. Vanbervliet, J. Fayette, C. Massacrier, C. Van Kooten, F. Brière, J. Banchereau and C. Caux. Dendritic cells enhance growth and differentiation of CD40-activated B lymphocytes. *J. Exp. Med.,* (1997), **185**, 941–951.
15. M. Matsumoto, Y. Fu, H. Molina, G. Huang, J. Kim, D. A. Thomas, M. H. Nahm and D. D. Chaplin. Distinct roles of lymphotoxin α and type 1 TNF receptor in the establishment of follicular dendritic cells from non-bone marrow-derived cells. *J. Exp. Med.*, (1997), **186**, 1997–2004.
16. F. Osorio and C. R. e Sousa. Myeloid C-type lectin receptors in pathogen recognition and host defense. *Immunity,* (2011), **34**, 651-664.
17. A. Alloatti, F. Kotsias, J. G. Magalhaes and S. Amigorena. Dendritic cell maturation and cross-presentation: timing matters! 2016. *Immunol. Rev.,* (2016), **272** (1), 97-108.
18. M. B. Lutz. Induction of CD4+ Regulatory and Polarized Effector/helper T Cells by Dendritic Cells. *Immune Netw.,* (2016), **16**, 13-25.
19. G. J. Adema, F. Hartgers, R. Verstraten, E. de Vries, G. Marland, S. Menon, J. Foster, Y. Xu, P. Nooyen, T. McClanahan, K. B. Bacon, and C. G. Figdor. A dendritic-cell-derived C-C chemokine that preferentially attracts naive T cells. *Nature*, (1997), **387**, 713–717.
20. A. Granelli-Piperno, M. Pope, K. Inaba, and R. M. Steinman. Coexpression of NF-kappa B/REL and SP1 transcription factors in HIV-1 induced, dendritic cell-T cell syncytia. *Proc. Natl Acad. Sci. USA,* (1995), **92**, 10944–10948.
21. I. Mellman, S. J. Turley and R. M. Steinman. Antigen processing for amateurs and professionals. *Trends Cell Biol.,* (1998), **8**, 231-237.
22. A. Chow, D. Toomre, W. Garrett and I. Mellman. Dendritic cell maturation triggers retrograde MHC class II transport from lysosomes to the plasma membrane. *Nature*, (2002), **418**, 988-994.
23. S. Kudo, K. Matsuno, T. Ezaki and M. Ogawa. A novel migration pathway for rat dendritic cells from the blood: Hepatic sinusoids-lymph translocation. *J. Exp. Med.,* (1997), **185**, 777–784.
24. R. M. Steinman, S. Turley, I. Mellman and K. Inaba. The induction of tolerance by dendritic cells that have captured apoptotic cells. *J. Exp. Med.,* (2000), **191**, 411–416.
25. T. Zal, A. Volkmann and B. Stockinger. Mechanisms of tolerance induction in major histocompatibility complex class II-restricted T cells specific for a blood-borne self-antigen. *J. Exp. Med.,* (1994), **180**, 2089–2099.
26. K. Inaba, M. Pack, M. Inaba, H. Sakuta, F. Isdell and R. M. Steinman. High levels of a major histocompatibility complex II–self peptide complex on dendritic cells from lymph node. *J. Exp. Med.,* (1997), **186**, 665–672.

27. J. M. Blander and R. Medzhitov. Toll-dependent selection of microbial antigens for presentation by dendritic cells. *Nature,* (2006), **440**, 808‑812.
28. A. Pichlmair and C. R. de Sousa. Innate recognition of viruses. *Immunity,* (2007), **27**, 370-383.
29. E. M. Creagh and L. A. O'Neill. TLRs, NLRs and RLRs: a trinity of pathogen sensors that co-operate in innate immunity. *Trends Immunol.,* (2006), **27**, 352–357.
30. C. Galustian, C. G. Park, W. Chai, M. Kiso, S. A. Bruening, Y.-S. Kang, R. M. Steinman and T. Feizi. High and low affinity carbohydrate ligands revealed for murine SIGN-R1 by carbohydrate array and cell binding approaches, and differing specificities for SIGN-R3 and Langerin. *Int. Immunol.,* (2004), **16**, 853-866.
31. M. E. Taylor and K. Drickamer. Convergent and divergent mechanisms of sugar recognition across kingdoms. *Curr. Opin. Struct. Biol.,* (2014), **28**, 14-22.
32. R. T. Lee, T.-L. Hsu, S. K. Huang, S.-L. Hsieh, C.-H. Wong and Y. C. Lee. Survey of immune-related, mannose/fucose-binding C-type lectin receptors reveals widely divergent sugar-binding specificities. *Glycobiology,* (2011), **21**, 512–520.
33. S. J. van Vliet, J. J. García-Vallejo and Y. van Kooyk. Dendritic cells and C-type lectin receptors: coupling innate to adaptive immune responses. *Immunol. Cell Biol.,* (2008), **86**, 580-587.
34. D. G. Nguyen and J. E. K. Hildreth. Involvement of macrophage mannose receptor in the binding and transmission of HIV by macrophages. *Eur. J. Immunol.,* (2003), **33**, 483–493.
35. T. Nagamatsu and D. J. Schust. The contribution of macrophages to normal and pathological pregnancies. *Am. J. Reprod. Immunol.,* (2010), **63**, 460–471.
36. C. G. Figdor, Y. van Kooyk and G. J. Adema. C-type lectin receptors on dendritic cells and Langerhans cells. *Nature Rev. Immunol.,* (2002), **2**, 77–84.
37. S. J. van Vliet, E. van Liempt, T. B. Geijtenbeek and Y. van Kooyk. Differential regulation of C-type lectin expression on tolerogenic dendritic cell subsets. *Immunobiology*, (2006), **211**, 577–585.
38. J. Valladeau, O. Ravel, C. Dezutter-Dambuyant, K. Moore, M. Kleijmeer, Y. Liu, V. Duvert-Frances, C. Vincent, D. Schmitt, J. Davoust, C. Caux, S. Lebecque and S. Saeland. Langerin, a novel C-type lectin specific to Langerhans cells, is an endocytic receptor that induces formation of Birbeck granules. *Immunity,* (2000), **12**, 71-81.
39. G. D. Brown. Dectin-1: a signalling non-TLR pattern-recognition receptor. *Nat. Rev. Immunol.,* (2006), **6**, 33–43.
40. M. P. Bergman, A. Engering, H. H. Smits, S. J. van Vliet, A. A. van Bodegraven, H. P. Wirth, M. L. Kapsenberg, C. M. J. E. Vandenbroucke-Grauls, Y. van Kooyk and B. J. Appelmelk. *Helicobacter pylori* modulates the T helper cell 1/T helper cell 2 balance through phase-variable interaction between lipopolysaccharide and DC-SIGN. *J. Exp. Med.,* (2004), **200**, 979–990.
41. A. P. West, A. A. Koblansky and S. Ghosh. Recognition and Signaling by Toll-Like Receptors. *Annu. Rev. Cell Dev. Biol.,* (2006), **22**, 409–437.

42. S. J. van Vliet, S. I. Gringhuis, T. B. Geijtenbeek and Y. van Kooyk. Regulation of effector T cells by antigen-presenting cells via interaction of the C-type lectin MGL with CD45. *Nat. Immunol.,* (2006), **7**, 1200–1208.
43. N. Kanazawa, T. Okazaki, H. Nishimura, K. Tashiro, K. Inaba and Y. Miyachi. DCIR acts as an inhibitory receptor depending on its immunoreceptor tyrosine-based inhibitory motif. *J. Invest. Dermatol.,* (2002), **118**, 261–266.
44. J. J. Garcia-Vallejo, E. van Liempt, M. P. da Costa, C. Beckers, B. van het Hof, S. I. Gringhuis, J.-J. Zwaginga, W. van Dijk, T. B. Geijtenbeek, Y. van Kooyk and I. van Die. DC-SIGN mediates adhesion and rolling of dendritic cells on primary human umbilical vein endothelial cells through LewisY antigen expressed on ICAM-2. *Mol. Immunol.,* (2008), **45**, 2359–2369.
45. V. Bogoevska, P. Nollau, L. Lucka, D. Grunow, B. Klampe, L. M. Uotila, A. Samsen, C. G. Gahmberg and C. Wagener. DC-SIGN binds ICAM-3 isolated from peripheral human leukocytes through Lewis X residues. *Glycobiology,* (2007), **17**, 324–333.
46. B. N. Gantner, R. M. Simmons, S. J. Canavera, S. Akira and D. M. Underhill. Collaborative induction of inflammatory responses by dectin-1 and Toll-like receptor 2. *J. Exp. Med.,* (2003), **197**, 1107–1117.
47. S. I. Gringhuis, J. den Dunnen, M. Litjens, B. van het Hof, Y. van Kooyk and T. B. Geijtenbeek. C-type lectin DC-SIGN modulates toll-like receptor signaling via Raf-1 kinase-dependent acetylation of transcription factor NF-kappa B. *Immunity*, (2007), **26**, 605–616.
48. J. S. Bonifacino and L. M. Traub. Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu. Rev. Biochem.,* (2003), **72**, 395–447.
49. T. B. Geijtenbeek and S. I. Gringhuis. Signalling through C-type lectin receptors: shaping immune responses. *Nat. Rev. Immunol.*, (2009), **9**, 465–479.
50. J. den Dunnen, S. I. Gringhuis and T. B. Geijtenbeek. Innate signaling by the C-type lectin DC-SIGN dictates immune responses. *Cancer Immunol. Immunother.,* (2009), **58**, 1149–57.
51. B. M. Curtis, S. Scharnowske and A. J. Watson. Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. *Proc. Natl. Acad. Sci. U S A,* (1992), **89**, 8356-8360.
52. J. J. García-Vallejo and Y. van Kooyk. The physiological role of DC-SIGN: a tale of mice and men. *Trends Immunol.,* (2013), **34**, 482-486.
53. T. B. Geijtenbeek and S. I. Gringhuis. C-type lectin receptors in the control of T helper cell differentiation. *Nat. Rev. Immunol.,* (2016), **16**, 433-448.
54. A. Cambi, K. Gijzen, I. J. de Vries, R. Torensma, B. Joosten, G. J. Adema, M. G. Netea, B. J. Kullberg, L. Romani and C. G. Figdor. The C-type lectin DC-SIGN (CD209) is an antigen-uptake receptor for *Candida albicans* on dendritic cells. *Eur. J. Immunol.,* (2003), **33**, 532–538.
55. T. B. Geijtenbeek, R. Torensma, S. J. van Vliet, G. C. F. van Duijnhoven, G. J. Adema, Y. van Kooyk and C. G. Figdor. Identification of DC-SIGN, a novel

dendritic cell–specific ICAM-3 receptor that supports primary immune responses. *Cell,* (2000), **100**, 575-585.

56. S. I. Gringhuis, J. den Dunnen, M. Litjens, M. van der Vlist and T. B. Geijtenbeek. Carbohydrate-specific signaling through the DC-SIGN signalosome tailors immunity to *Mycobacterium tuberculosis*, HIV-1 and *Helicobacter pylori*. *Nat. Immunol.*, (2009), **10**, 1081–1088.
57. M. Okano, A. R. Satoskar, K. Nishizaki, M. Abe and D. A. Harn Jr. Induction of TH2 responses and IgE is largely due to carbohydrates functioning as adjuvants on *Schistosoma mansoni* egg antigens. *J. Immunol.*, (1999), **163**, 6712–6717.
58. S. I. Gringhuis, T. M. Kaptein, B. A. Wevers, M. van der Vlist, E. J. Klaver, I. van Die, L. E. M. Vriend, M. A. W. P. de Jonj and T. B. Geijtenbeek. Fucose-based PAMPs prime dendritic cells for follicular T helper cell polarization via DC-SIGN-dependent IL-27 production. *Nat. Commun.,* (2014), **5,** 5074.
59. T. B. Geijtenbeek, D. J. E. B. Krooshoop, D. A. Bleijs, S. J. van Vliet, G. C. F. van Duijnhoven, V. Grabovsky, R. Alon, C. G. Figdor and Y. van Kooyk. DC-SIGN–ICAM-2 interaction mediates dendritic cell trafficking. *Nat. Immunol.,* (2000), **1**, 353-357.
60. K. P. J. M. van Gisbergen, M. Sanchez-Hernandez, T. B. Geijtenbeek and Y. van Kooyk. Neutrophils mediate immune modulation of dendritic cells through glycosylation-dependent interactions between Mac-1 and DC-SIGN. *J. Exp. Med.,* (2005), **201**, 1281–1292.
61. G. D. Brown, J. A. Willment and L. Whitehead. C-type lectins in immunity and homeostasis. *Nat. Rev. Immunol.,* (2018), **18**, 374-389.
62. N. Rieber, A. Singh, H. Öz, M. Carevic, M. Bouzani, J. Amich, M. Ost, Z. Ye, M. Ballbach, I. Schäfer, M. Mezger, S. N. Klimosch, A. N. R. Weber, R. Handgretinger, S. Krappmann, J. Liese, M. Engeholm, R. Schüle, H. R. Salih, L. Marodi, C. Speckmann, B. Grimbacher, J. Ruland, G. D. Brown, A. Beilhack, J. Loeffler and D. Hartl. Pathogenic fungi regulate immunity by inducing neutrophilic myeloid-derived suppressor cells. *Cell Host Microbe,* (2015), **17**, 507–514.
63. F. V. Loures, M. Röhm, C. K. Lee, E. Santos, J. P. Wang, C. A. Specht, V. L. G. Calich, C. F. Urban and S. M. Levitz. Recognition of *Aspergillus fumigatus* hyphae by human plasmacytoid dendritic cells is mediated by dectin-2 and results in formation of extracellular traps. *PLoS Pathog.,* (2015), **11**, e1004643.
64. A. Vázquez, J.-D. Ruiz-Rosado, L. I. Terrazas, I. Juárez, L. Gomez-Garcia, E. Calleja, G. Camacho, A. Chávez, M. Romero, T. Rodriguez, B. Espinoza and M. Rodriguez-Sosa. Mouse macrophage galactose-type lectin (mMGL) is critical for host resistance against *Trypanosoma cruzi* infection. *Int. J. Biol. Sci.,* (2014), **10**, 909–920.
65. G. J. Wilson, M. J. Marakalala, J. C. Hoving, A. Laarhoven, R. A. Drummond, B. Kerscher, R. Keeton, E. de Vosse, T. H. M. Ottenhoff, T. S. Plantinga, B. Alisjahbana, D. Govender, G. S. Besra, M. G. Netea, D. M. Reid, J. A. Willment, M. Jacobs, S. Yamasaki, R. van Crevel and G. D. Brown. The C-type lectin

receptor CLECSF8/CLEC4D is a key component of anti-mycobacterial immunity. *Cell Host Microbe,* (2015), **17**, 252–259.

66. F. Behler-Janbeck, T. Takano, R. Maus, J. Stolper, D. Jonigk, M. T. Tarrés, T. Fuehner, A. Prasse, T. Welte, M. S. M. Timmer, B. L. Stocker, Y. Nakanishi, T. Miyamoto, S. Yamasaki and U. A. Maus. C-type lectin Mincle recognizes glucosyl-diacylglycerol of *Streptococcus pneumoniae* and plays a protective role in pneumococcal pneumonia. *PLoS Pathog.,* (2016), **12**, e1006038.
67. A. Sharma, A. L. Steichen, C. N. Jondle, B. B. Mishra and J. Sharma. Protective role of Mincle in bacterial pneumonia by regulation of neutrophil mediated phagocytosis and extracellular trap formation. *J. Infect. Dis.,* (2014), **209**, 1837–1846.
68. L. de Witte, A. Nabatov, M. Pion, D. Fluitsma, M. A. de Jong, T. de Gruijl, V. Piguet, Y van Kooyk and T. B. Geijtenbeek. Langerin is a natural barrier to HIV-1 transmission by Langerhans cells. *Nat. Med.,* (2007), **13**, 367–371.
69. Y. Kimura, A. Inoue, S. Hangai, S. Saijo, H. Negishi, J. Nishio, S. Yamasaki, Y. Iwakura, H. Yanai and T. Taniguchi. The innate immune receptor dectin-2 mediates the phagocytosis of cancer cells by Kupffer cells for the suppression of liver metastasis. *Proc. Natl. Acad. Sci. USA*, (2016), **113**, 14097–14102.
70. D. Ding, Y. Yao, S. Zhang, C. Su and Y. Zhang. C-type lectins facilitate tumor metastasis. *Oncol. Lett.,* (2017), **13**, 13–21.
71. J. J. Garcia-Vallejo and Y. van Kooyk. DC-SIGN: the strange case of Dr. Jekyll and Mr. Hyde. *Immunity,* (2015), **42**, 983-985.
72. J. T. Monteiro and B. Lepenies. Myeloid C-type lectin receptors in viral recognition and antiviral immunity. *Viruses,* (2017), **9**, E59.
73. C. P. Mason and A. W. Tarr. Human lectins and their roles in viral infections. *Molecules,* (2015), **20**, 2229-2271.
74. Y. van Kooyk and T. B. Geijtenbeek. DC-SIGN: escape mechanism for pathogens. *Nat. Rev. Immunol.,* (2003), **3**, 697-709.
75. R. Sarkar, D. Mitra and S. Chakrabarti. HIV-1 gp120 protein downregulates Nef induced IL-6 release in immature dentritic cells through interplay of DC-SIGN. *PLoS One,* (2013), **8**, e59073.
76. A. Pawlowski, M. Jansson, M. Sköld, M. E. Rottenberg and G. Källenius. Tuberculosis and HIV co-infection. *PloS ONE,* (2012), **8**, e1002464.
77. T. B. Geijtenbeek, D. S. Kwon, R. Torensma, S. J. van Vliet, G. C. van Duijnhoven, J. Middel, I. L. Cornelissen, H. S. Nottet, V. N. KewalRamani, D. R. Littman, C. G. Figdor and Y. van Kooyk. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell,* (2000), **100**, 587−597.
78. C. M. Ribeiro, R. Sarrami-Forooshani, L. C. Setiawan, E. M. Zijlstra-Willems, J. L. van Hamme, W. Tigchelaar, N. N. van der Wel, N. A. Kootstra, S. I. Gringhuis and T. B. Geijtenbeek. Receptor usage dictates HIV-1 restriction by human TRIM5alpha in dendritic cell subsets. *Nature,* (2016), **540**, 448–452.

79. D. S. Kwon, G. Gregorio, N. Bitton, W. A. Hendrickson and D. R. Littman. DC-SIGN-mediated internalization of HIV is required for trans-enhancement of T cell infection. *Immunity,* (2002), **16**, 135–144.

80. K. Hijazi, Y. Wang, C. Scala, S. Jeffs, C. Longstaff, D. Stieh, B. Haggarty, G. Vanham, D. Schols, J. Balzarini, I. M. Jones, J. Hoxie, R. Shattock and C. G. Kelly. DC-SIGN increases the affinity of HIV-1 envelope glycoprotein interaction with CD4. *PloS ONE,* (2011), **6**, e28307.

81. S. I. Gringhuis, M. van der Vlist, L. M. van den Berg, J. den Dunnen, M. Litjens and T. B. Geijtenbeek. HIV-1 exploits innate signaling by TLR8 and DC-SIGN for productive infection of dendritic cells. *Nat. Immunol.,* (2010), **11**, 419-426.

82. E. Navarro-Sanchez, R. Altmeyer, A. Amara, O. Schwartz, F. Fieschi, J. L. Virelizier, F. Arenzana-Seisdedos and P. Desprès. Dendritic-cell-specific ICAM3-grabbing non-integrin is essential for the productive infection of human dendritic cells by mosquito-cell-derived dengue viruses. *EMBO rep.,* (2003), **4**, 1–6.

83. B. Tassaneetrithep, T. H. Burgess, A. Granelli-Piperno, C. Trumpfheller, J. Finke, W. Sun, M. A. Eller, K. Pattanapanyasat, S. Sarasombath, D. L. Birx, R. M. Steinman, S. Schlesinger and M. A. Marovich. DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. *J. Exp. Med.,* (2003), **197**, 823–829.

84. E. Pokidysheva, Y. Zhang, A. J. Battisti, C. M. Bator-Kelly, P. R. Chipman, C. Xiao, G. G. Gregorio, W. A. Hendrickson, R. J. Kuhn and M. G. Rossmann. Cryo-EM reconstruction of dengue virus in complex with the carbohydrate recognition domain of DC-SIGN. *Cell,* (2006), **124**, 485-493.

85. P. Liu, M. Ridilla, P. Patel, L. Betts, E. Gallichotte, L. Shahidi, N. L. Thompson and K. Jacobson. Beyond attachment: Roles of DC-SIGN in dengue virus infection. *Traffic,* (2017), **18**, 218-231.

86. M. M. Alen, K. Dallmeier, J. Balzarini, J. Neyts and D. Schols. Crucial role of the N-glycans on the viral E-envelope glycoprotein in DC-SIGN-mediated dengue virus infection. *Antiviral Res.,* (2012), **96**, 280-287.

87. P. Y. Lozach, L. Burleigh, I. Staropoli, E. Navarro-Sanchez, J. Harriague, J. L. Virelizier, F. A. Rey, P. Desprès, F. Arenzana-Seisdedos and A. Amara. Dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin (DC-SIGN)-mediated enhancement of dengue virus infection is independent of DC-SIGN internalization signals. *J. Biol. Chem.,* (2005), **280**, 23698-23708.

88. G. Simmons, J. D. Reeves, C. C. Grogan, L. H. Vandenberghe, F. Baribaud, J. C. Whitbeck, E. Burke, M. J. Buchmeier, E. J. Soilleux, J. L. Riley, R. W. Doms, P. Bates and S. Pöhlmann. DC-SIGN and DC-SIGNR bind ebola glycoproteins and enhance infection of macrophages and endothelial cells. *Virology,* (2003), **305**, 115–123.

89. C. P. Alvarez, F. Lasala, J. Carrillo, O. Muñiz, A. L. Corbí and R. Delgado. C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus *in cis* and *in trans*. *J. Virol.,* (2002), **76**, 6841–6844.

90. N. Mulherkar, M. Raaben, J. C. de la Torre, S. P. Whelan and K. Chandran. The Ebola virus glycoprotein mediates entry via a non-classical dynamin-dependent macropinocytic pathway. *Virology,* (2011), **419**, 72-83.

91. J. E. Carette, M. Raaben, A. C. Wong, A. S. Herbert, G. Obernosterer, N. Mulherkar, A. I. Kuehne, P. J. Kranzusch, A. M. Griffin, G. Ruthel, P. Dal Cin, J. M. Dye, S. P. Whelan, K. Chandran and T. R. Brummelkamp. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature,* (2011), **477,** 340-343.

92. P. Aleksandrowicz, A. Marzi, N. Biedenkopf, N. Beimforde, S. Becker, T. Hoenen, H. Feldmann, and H. J. Schnittler. Ebola virus enters host cells by macropinocytosis and clathrin-mediated endocytosis. *J. Infect. Dis.,* (2011), **204**, S957-967.

93. F. Halary, A. Amara, H. Lortat-Jacob, M. Messerle, T. Delaunay, C. Houlès, F. Fieschi, F. Arenzana-Seisdedos, J. F. Moreau and J. Déchanet-Merville. Human cytomegalovirus binding to DC-SIGN is required for dendritic cell infection and target cell trans-infection. *Immunity,* (2002), **17**, 653–664 .

94. N. Plazolles, J. M. Humbert, L. Vachot, B. Verrier, C. Hocke and F. Halary. Pivotal advance: The promotion of soluble DC-SIGN release by inflammatory signals and its enhancement of cytomegalovirus-mediated cis-infection of myeloid dendritic cells. *J. Leukoc. Biol.,* (2011), **89**, 329-342.

95. S. Pöhlmann, J. Zhang, F. Baribaud, Z. Chen, G. J. Leslie, G. Lin, A. Granelli-Piperno, R. W. Doms, C. M. Rice and J. A. McKeating. Hepatitis C virus glycoproteins interact with DC-SIGN and DC-SIGNR. *J. Virol.,* (2003), **77**, 4070–4080.

96. P. Y. Lozach, H. Lortat-Jacob, A. de Lacroix de Lavalette, I. Staropoli, S. Foung, A. Amara, C. Houles, F. Fieschi, O. Schwartz, J. L. Virelizier, F. Arenzana-Seisdedos and R. Altmeyer. DC-SIGN and L-SIGN are high–affinity binding receptors for hepatitis C virus glycoprotein E2. *J. Biol. Chem.,* (2003), **278**, 20358–20366.

97. P. Wang, K. Hu, S. Luo, M. Zhang, X. Deng, C. Li, W. Jin, B. Hu, S. He, M. Li, T. Du, G. Xiao, B. Zhang, Y. Liu and Q. Hu. DC-SIGN as an attachment factor mediates Japanese encephalitis virus infection of human dendritic cells via interaction with a single high-mannose residue of viral E glycoprotein. *Virology,* (2016), **488**, 108-119.

98. I. Phoenix, S. Nishiyama, N. Lokugamage, T. E. Hill, M. B. Huante, O. A. Slack, V. H. Carpio, a. N. Freiberg and T. Ikegami. N-Glycans on the Rift Valley Fever Virus Envelope Glycoproteins Gn and Gc Redundantly Support Viral Infection via DC-SIGN. *Viruses,* (2016), **8**, E149.

99. N. Kerur, M. V. Veettil, N. Sharma-Walia, S. Sadagopan, V. Bottero, A. G. Paul and B. Chandran. Characterization of entry and infection of monocytic THP-1 cells by Kaposi's sarcoma associated herpesvirus (KSHV): role of heparan sulfate, DC-SIGN, integrins and signaling. *Virology,* (2010), **406**, 103-116.

100. a) L. de Witte, R. D. de Vries, M. van der Vlist, S. Yüksel, M. Litjens, R. L. de Swart and T. B. Geijtenbeek. DC-SIGN and CD150 have distinct roles in transmission of measles virus from dendritic cells to T-lymphocytes. *PLoS*

Pathog., (2008), **4**, e1000049. b) A. W. Mesman, E. M. Zijlstra-Willems, T. M. Kaptein, R. L. de Swart, M. E. Davis, M. Ludlow, W. P. Duprex, M. U. Gack, S. I. Gringhuis and T. B. Geijtenbeek. Measles virus suppresses RIG-like receptor activation in dendritic cells via DC-SIGN mediated inhibition of PP1 phosphatases. *Cell Host Microbe,* (2014), **16**, 31–42.

101. P. Léger, M. Tetard, B. Youness, N. Cordes, R. N. Rouxel, M. Flamand and P. Y. Lozach. Differential use of the C-Type lectins L-SIGN and DC-SIGN for Phlebovirus endocytosis. *Traffic,* (2016), **17**, 639-656.
102. L. Gillespie, P. Roosendahl, W. C. Ng, A. G. Brooks, P. C. Reading and S. L. Londrigan. Endocytic function is critical for influenza A virus infection via DC-SIGN and L-SIGN. *Sci. Rep.,* (2016), **6**, 19428.
103. L. Tailleux, O. Schwartz, J. L. Herrmann, E. Pivert, M. Jackson, A. Amara, L. Legres, D. Dreher, L. P. Nicod, J. C. Gluckman, P. H. Lagrange, B. Gicquel and O. Neyrolles. DC-SIGN is the major *Mycobacterium tuberculosis* receptor on human dendritic cells*. J. Exp. Med.,* (2003), **197**, 121–127.
104. S. Ehlers. DC-SIGN and mannosylated surface structures of *Mycobacterium tuberculosis*: a deceptive liaison. *Eur. J. Cell. Biol.,* (2010), **89**, 95-101.
105. Y. van Kooyk, B. Appelmelk and T. B. Geijtenbeek. A fatal attraction: *Mycobacterium tuberculosis* and HIV-1 target DC-SIGN to escape immune surveillance. *Trends Mol. Med.,* (2003), **9**, 153-159.
106. I. Vergne, M. Gilleron and J. Nigou. Manipulation of the endocytic pathway and phagocyte functions by *Mycobacterium tuberculosis* lipoarabinomannan. *Front. Cell Infect. Microbiol.,* (2015), **4**, 187.
107. T. B. Geijtenbeek, S. J. Van Vliet, E. A. Koppel, M. Sanchez-Hernandez, C. M. Vandenbroucke-Grauls, B. Appelmelk and Y. Van Kooyk. *Mycobacteria* target DC-SIGN to suppress dendritic cell function. *J. Exp. Med.,* (2003), **197**, 7–17.
108. L. Balboa, M. M. Romero, N. Yokobori, P. Schierloh, l. Geffner, J. I. Basile, R. M. Musella, E. Abbate, S. de la Barrera, M. C. Sasiain and M. Alemán. *Mycobacterium tuberculosis* impairs dendritic cell response by altering CD1b, DC-SIGN and MR profile. *Immunol. Cell. Biol.,* (2010), **88**, 716-726.
109. A. R. El-Awady, B. Miles, E. Scisci, Z. B. Kurago, C. D. Palani, R. M. Arce, J. L. Waller, C. A. Genco, C. Slocum, M. Manning, P. V. Schoenlein and C. W. Cutler. *Porphyromonas gingivalis* evasion of autophagy and intracellular killing by human myeloid dendritic cells involves DC-SIGN-TLR2 crosstalk. *PLoS Pathog.,* (2015), **10**, e1004647.
110. N. A. Gow, F. L. van de Veerdonk, A. J. Brown and M. G. Netea. *Candida albicans* morphogenesis and host defence: discriminating invasion from colonization. *Nat. Rev. Microbiol., (*2011), **10**, 112-122.
111. G. G. Gauglitz, H. Callenberg, G. Weindl and H. C. Korting. Host defence against *Candida albicans* and the role of pattern-recognition receptors. *Acta Derm. Venereol.,* (2012), **92**, 291-298.
112. A. Cambi, M. G. Netea, H. M. Mora-Montes, N. A. R. Gow, S. V. Hato, D. W. Lowman, B.-J. Kullberg, R. Torensma, D. L. Williams and C. G. Figdor. Dendritic

cell interaction with *Candida albicans* critically depends on N-linked mannan. *J. Biol. Chem.,* (2008), **283**, 20590-20599.

113. S. Mari, D. Serrano-Gómez, F. J. Cañada, A. L. Corbí and J. Jiménez-Barbero. 1D Saturation Transfer Difference NMR experiments on living cells: the DC-SIGN/oligomannose interaction. *Angew. Chem. Int. Ed.,* (2005), **44**, 296 –298.
114. R. P. da Silva, C. Heiss, I. Black, P. Azadi, J. Q. Gerlach, L. R. Travassos, L. Joshi, M. Kilcoyne and R. Puccia. Extracellular vesicles from *Paracoccidioides* pathogenic species transport polysaccharide and expose ligands for DC-SIGN receptors. *Sci. Rep.,* (2015), **5**, 14213.
115. I. van Die, S. J. van Vliet, A. K. Nyame, R. D. Cummings, C. M.C. Bank, B. Appelmelk, T. B. Geijtenbeek and Y. van Kooyk. The dendritic cell-specific C-type lectin DC-SIGN is a receptor for *Schistosoma mansoni* egg antigens and recognizes the glycan antigen Lewis X. *Glycobiology,* (2003), **13**, 471-478.
116. M. Colmenares, A. Puig-Kroger, O. M. Pello, A. L. Corbi and L. Rivas. Dendritic cell (DC)-specific intercellular adhesion molecule 3 (ICAM-3)-grabbing nonintegrin (DC-SIGN, CD209), a C-type surface lectin in human DCs, is a receptor for *Leishmania* amastigotes. *J. Biol. Chem.,* (2002), **277**, 36766–36769.
117. E. Caparrós, D. Serrano, A. Puig-Kröger, L. Riol, F. Lasala, I. Martinez, F. Vidal-Vanaclocha, R. Delgado, J. L. Rodríguez-Fernández, L. Rivas, A. L. Corbí and M. Colmenares. Role of the C-type lectins DC-SIGN and L-SIGN in *Leishmania* interaction with host phagocytes. *Immunobiology,* (2005), **210**, 185–193.
118. C. A. Carlos, H. F. Dong, O. M. Howard, J. J. Oppenheim , F. G. Hanisch and O. J. Finn. Human tumor antigen MUC1 is chemotactic for immature dendritic cells and elicits maturation but does not promote Th1 type immunity. *J. Immunol.,* (2005), **175**, 1628–1635.
119. C. Napoletano, A. Rughetti, M. P. A. Tarp, J. Coleman, E. P. Bennett, G. Picco, P. Sale, K. Denda-Nagai, T. Irimura, U. Mandel, H. Clausen, L. Frati, J. Taylor-Papadimitriou, J. Burchell and M. Nuti. Tumor-associated Tn-MUC1 glycoform is internalized through the macrophage galactose-type C-type lectin and delivered to the HLA class I and II compartments in dendritic cells. *Cancer Res.,* (2007), **67**, 8358–8367.
120. M. Nonaka, B. Y. Ma, R. Murai, N. Nakamura, M. Baba, N. Kawasaki, K. Hodohara, S. Asano and T. Kawasaki. Glycosylation dependent interactions of C-type lectin DC-SIGN with colorectal tumor-associated Lewis glycans impair the function and differentiation of monocyte-derived dendritic cells. *J. Immunol.,* (2008), **180**, 3347–3356.
121. Y. Zhang, Y. Luo, W. Li, J. Liu, M. Chen, H. Gu, B. Wang and X. Yao. DC-SIGN promotes allergen uptake and activation of dendritic cells in patients with atopic dermatitis. *J. Dermatol. Sci.,* (2016), **84**, 128-136.
122. E. Axelgaard, J. A. Ostergaard, S. Thiel and T. K. Hansen. Diabetes is associated with increased autoreactivity of mannan-binding lectin. *J. Diabetes Res.,* (2017), 6368780.

123. T. Uto, T. Fukaya, H. Takagi, K. Arimura, T. Nakamura, N. Kojima, B. Malissen and K. Sato. Clec4A4 is a regulatory receptor for dendritic cells that impairs inflammation and T-cell immunity. *Nat. Commun.,* (2016), **7**, 11273.

124. M. Kamalakannan, L. M. Chang, G. Grishina, H. A. Sampson and M. Masilamani. Identification and characterization of DC-SIGN-binding glycoproteins in allergenic foods. *Allergy,* (2016), **71**, 1145–1155.

125. F. Salazar, L. Hall, O. H. Negm, D. Awuah, P. J. Tighe, F. Shakib and A. M. Ghaemmaghami. The mannose receptor negatively modulates the Toll-like receptor 4-aryl hydrocarbon receptor-indoleamine 2,3-dioxygenase axis in dendritic cells affecting T helper cell polarization. *J. Allergy Clin. Immunol.,* (2016), **137**, 1841–1851.e2.

126. a) K. Drickamer. Demonstration of carbohydrate recognition activity in diverse proteins which share a common primary structure motif. *Biochem. Soc. Trans.,* (1989), 17, 13‑15. b) K. Drickamer. Two distinct classes of carbohydrate-recognition domains in animals lectins. *J. Biol. Chem.,* (1988), 263, 9557-9560.

127. N. Matsumoto, R. K. Ribaudo, J. P. Abastado, D. H. Margulies and W. M. Yokoyama. The lectin-like NK cell receptor Ly-49A recognizes a carbohydrate-independent epitope on its MHC class I ligand. *Immunity,* (1998), **8**, 245‑254.

128. G. D. Brown and S. Gordon. Immune recognition. A new receptor for beta-glucans. *Nature,* (2001), **413**, 36‑37.

129. M. E. Taylor and K. Drickamer. C-type lectin family: Overview. In *Glycoscience: Biology and Medicine*, eds. N. Taniguchi, T. Endo, G. W. Hart, P. H. Seeberger and C.-H. Wong. Springer Verlag (Japan), 1st edn, 2015, **ch. 125**.

130. A. N. Zelensky and J. E. Gready. The C-type lectin-like domain superfamily. *FEBS J.*, (2005), **272**, 6179-6217.

131. R. D. Cummings and R. P. McEver. C-type lectins. In *Essentials of Glycobiology*, eds. A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar and P. H. Seeberger. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press (New York), 3rd edn, 2015-2017, **ch. 34**.

132. A. N. Zelensky and J. E. Gready. Comparative analysis of structural properties of the C-type-lectin-like domain (CTLD). *Proteins,* (2003), **52**, 466‑477.

133. O. Lysenko, D. Schulte, M. Mittelbronn and A. Steinle. BACL is a novel brain-associated, non-NKC-encoded mammalian C-Type lectin-like receptor of the CLEC2 family. *PloS ONE,* (2013), **8**, e65345.

134. S. Bhalchandra, J. Ludington, I. Coppens and H. D. Ward. Identification and characterization of *Cryptosporidium parvum* CLEC, a novel C-Type lectin domain-containing Mucin-like glycoprotein. *Infect. Immun.,* (2013), **81**, 3356-3365.

135. K. Drickamer. C-type lectin-like domains. *Curr. Opin. Struct. Biol.,* (1999), **9**, 585–590.

136. A. Furukawa, J. Kamishikiryo, D. Mori, K. Toyonaga, Y. Okabe, A. Toji, R. Kanda,Y. Miyake, T. Ose, S. Yamasaki and K. Maenaka. Structural analysis for

glycolipid recognition by the C-type lectins Mincle and MCL. *Proc. Natl. Acad. Sci.,* (2013), **110**, 17438–17443.

137. W. I. Weis and K. Drickamer. Structural basis of lectin-carbohydrate interaction. *Annu. Rev. Biochem.*, (1996), **65**, 441-473.
138. A. K. Shrive, H. A. Tharia, P. Strong, U. Kishore, I. Burns, P. J. Rizkallah, K. B. Reid and T. J. Greenhough. High resolution structural insights into ligand binding and immune cell recognition by human lung surfactant protein D. *J. Mol. Biol.,* (2003), **331**, 509‑523.
139. J. A. Loeb and K. Drickamer. Conformational changes in the chicken receptor for endocytosis of glycoproteins. Modulation of ligand-binding activity by Ca^{2+} and pH. *J. Biol. Chem.,* (1988), **263**, 9752-9760.
140. T. Onizuka, H. Shimizu, Y. Moriwaki, T. Nakano, S. Kanai, I. Shimada and H. Takahashi. NMR study of ligand release from asialoglycoprotein receptor under solution conditions in early endosomes. *FEBS J.,* (2012), **279**, 2645–2656.
141. T. B. Geijtenbeek, G. C. van Duijnhoven, S. J. van Vliet, E. Krieger, G. Vriend, C. G. Figdor and Y. van Kooyk. Identification of different binding sites in the dendritic cell-specific receptor DC-SIGN for intercellular adhesion molecule 3 and HIV-1. *J. Biol. Chem.,* (2002), **277**, 11314-11320.
142. M. Meier, M. D. Bider, V. N. Malashkevich, M. Spiess and P. Burkhard. Crystal structure of the carbohydrate recognition domain of the H1 subunit of the asialoglycoprotein receptor. *J. Mol. Biol.,* (2000), **300**, 857‑865.
143. B. G. Keller and C. Rademacher. Allostery in C-type lectins. *Curr. Opin. Struct. Biol.,* (2019), **62**, 31-38.
144. J. H. Graversen, R. H. Lorentsen, C. Jacobsen, S. K. Moestrup, B. W. Sigurskjold, H. C. Thogersen and M. Etzerodt. The plasminogen binding site of the C-type lectin tetranectin is located in the carbohydrate recognition domain, and binding is sensitive to both calcium and lysine. *J. Biol. Chem.,* (1998), **273**, 29241-29246.
145. S. Wragg and K. Drickamer. Identification of amino acid residues that determine pH dependence of ligand binding to the asialoglycoprotein receptor during endocytosis. *J. Biol. Chem.,* (1999), **274**, 35400‑35406.
146. J. Hanske, S. Aleksić, M. Ballaschk, M. Jurk, E. Shanina, M. Beerbaum, P. Schmieder, B. G. Keller and C. Rademacher. Intradomain allosteric network modulates calcium affinity of the C-type lectin receptor Langerin. *J. Am. Chem. Soc.*, (2016), 138, 12176-12186.
147. K. K. Ng, S. Park-Snyder and W. I. Weis. Ca^{2+}-dependent structural changes in C-type mannose-binding proteins. *Biochemistry,* (1998), **37**, 17965-17976.
148. M. E. Taylor and K. Drickamer. *Introduction to Glycobiology*. Oxford University Press (Oxford), 3rd edn, 2003.
149. N. S. Stambach and M. E. Taylor. Characterization of carbohydrate recognition by Langerin, a C-type lectin of Langerhans cells. *Glycobiology,* (2003), **13**, 401-410.

150. W. Liu, L. Tang, G. Zhang, H. Wei, Y. Cui, L. Guo, Z. Gou, X. Chen, D. Jiang, Y. Zhu, G. Kang and F. He. Characterization of a novel C-type lectin-like gene, LSECtin: demonstration of carbohydrate binding and expression in sinusoidal endothelial cells of liver and lymph node. *J. Biol. Chem.,* (2004), **279**, 18748‑18758.

151. S. J. van Vliet, E. Saeland and Y. van Kooyk. Sweet preferences of MGL: carbohydrate specificity and function. *Trends Immunol.,* (2008), **29**, 83-90.

152. S. J. van Vliet, E. van Liempt, E. Saeland, C. A. Aarnoudse, B. Appelmelk, T. Irimura, T. B. Geijtenbeek, O. Blixt, R. Alvarez, I. van Die and Y. van Kooyk. Carbohydrate profiling reveals a distinctive role for the C-type lectin MGL in the recognition of helminth parasites and tumor antigens by dendritic cells. *Int. Immunol.,* (2005), **17**, 661–669.

153. R. J. Stockert. The asialoglycoprotein receptor: relationships between structure, function, and expression. *Physiol. Rev.,* (1995), **75**, 591-609.

154. T. Yoshida, Y. Tsuruta, M. Iwasaki, S. Yamane, T. Ochi and R. Suzuki. SRCL/CL-P1 recognizes GalNAc and a carcinoma-associated antigen, Tn antigen. *J. Biochem. (Tokyo),* (2003), **133**, 271-277.

155. P. J. Coombs, S. A. Graham, K. Drickamer and M. E. Taylor. Selective binding of the scavenger receptor Ctype lectin to LewisX trisaccharide and related glycan ligands. *J. Biol. Chem.,* (2005), **280**, 22993‑22999.

156. D. A. Mitchell, A. J. Fadden and K. Drickamer. A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR. Subunit organization and binding to multivalent ligands. *J. Biol. Chem.,* (2001), **276**, 28939-28945.

157. K. K. Ng, K. Drickamer and W. I. Weis. Structural analysis of monosaccharide recognition by rat liver mannose-binding protein. *J. Biol. Chem.,* (1996), **271**, 663-674.

158. S. Elgavish and B. Shaanan. Lectin-carbohydrate interactions: different folds, common recognition principles. *Trends Biochem. Sci.,* (1997), **22**, 462‑467.

159. a) S. F. Poget, G. B. Legge, M. R. Proctor, P. J. Butler, M. Bycroft and R. L. Williams. The structure of a tunicate C-type lectin from *Polyandrocarpa misakiensis* complexed with D-galactose. *J. Mol. Biol.,* (1999), **290**, 867-879. b) A. Imberty, C. Gautier, J. Lescar, S. Pérez, L. Wyns and R. J. Loris. An unusual carbohydrate binding site revealed by the structures of two *Maackia amurensis* lectins complexed with sialic acid-containing oligosaccharides. *J. Biol. Chem.,* (2000), **275**, 17541-17548.

160. A. R. Kolatkar and W. I. Weis. Structural basis of galactose recognition by C-type animal lectins. *J. Biol. Chem.,* (1996), **271**, 6679-6685.

161. K. K. Ng, A. R. Kolatkar, S. Park-Snyder, H. Feinberg, D. A. Clark, K. Drickamer and W. I. Weis. Orientation of bound ligands in mannose-binding proteins. Implications for multivalent ligand recognition. *J. Biol. Chem.,* (2002), **277**, 16088‑16095.

162. M. Rinnbauer, B. Ernst, B. Wagner, J. Magnani, A. J. Benie and T. Peters. Epitope mapping of sialyl LewisX bound to E-selectin using saturation transfer difference NMR experiments. *Glycobiology,* (2003), **13**, 435-443.
163. H. Tateno, K. Ohnishi, R. Yabe, N. Hayatsu, T. Sato, M. Takeya, H. Narimatsu and J. Hirabayashi. Dual specificity of Langerin to sulfated and mannosylated glycans via a single C-type carbohydrate recognition domain. *J. Biol. Chem.,* (2010), **285**, 6390-6400.
164. W. S. Somers, J. Tang, G. D. Shaw and R. T. Camphausen. Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1. *Cell,* (2001), **105**, 957–969.
165. H. Feinberg, S. A. F. Jégouzo, T. J. W. Rowntree, Y. Guan, M. A. Brash, M. E. Taylor, W. I. Weis and K. Drickamer. Mechanism for recognition of an unusual mycobacterial glycolipid by the macrophage receptor Mincle. *J. Biol. Chem.,* (2013), **288**, 28457-28465.
166. A. S. Llera, F. Viedma, F. Sánchez-Madrid and J. Tormo. Crystal structure of the C-type lectin-like domain from the human hematopoietic cell receptor CD69. *J. Biol. Chem., (2000),* **276**, 7312-7319.
167. J. Brown, C.A. O'Callaghan, A. S. Marshall, R. J. Gilbert, C. Siebold, S. Gordon, G. D. Brown and E. Y. Jones. Structure of the fungal beta-glucan-binding immune receptor dectin-1: implications for function. *Protein Sci.,* (2007), **16**, 1042-1052.
168. F. Teillet, B. Dublet, J.-P. Andrieu, C. Gaboriaud, G. J. Arlaud and N. M. Thielens. The two major oligomeric forms of human mannan-binding lectin: chemical characterization, carbohydrate-binding properties, and interaction with MBL-associated serine proteases. *J. Immunol.,* (2005), **174**, 2870-2877.
169. S. A. F. Jégouzo, A. Quintero-Martínez, X. Ouyang, Á. dos Santos, M. E. Taylor and K. Drickamer. Organization of the extracellular portion of the macrophage galactose receptor: A trimeric cluster of simple binding sites for N-acetylgalactosamine. *Glycobiology,* (2013), **23**, 853–864.
170. H. Feinberg, C. K. W. Tso, M. E. Taylor, K. Drickamer and W. I. Weis. Segmented helical structure of the neck region of the glycan-binding receptor DC-SIGNR. *J. Mol. Biol.,* (2009), **394**, 613-620.
171. G. A. Snyder, M. Colonna and P. D. Sun. The structure of DC-SIGNR with a portion of its repeat domain lends insights to modeling of the receptor tetramer. *J. Mol. Biol.,* (2005), **347**, 979-989.
172. E. J. Soilleux. DC-SIGN (dendritic cell-specific ICAM-grabbing non-integrin) and DC-SIGN-related (DC-SIGNR): friend or foe. *Clin. Sci. (Lond),* (2003), **104**, 437-446.
173. J. F. Head, T. R. Mealy, F. X. McCormack and B. A. Seaton. Crystal structure of trimeric carbohydrate recognition and neck domains of surfactant protein A. *J. Biol. Chem.,* (2003), **278**, 43254–43260.
174. R. Wallis, J. M. Shaw, J. Uitdehaag, C.-B. Chen, D. Torgersen and K. Drickamer. Localization of the serine protease-binding sites in the collagen-like domain of mannose-binding protein. *J. Biol. Chem.,* (2004), **278**, 14065–14073.

175. S. Menon, K. Rosenberg, S. A. Graham, E. M. Ward, M. E. Taylor, K. Drickamer and D. E. Leckband. Binding-site geometry and flexibility in DC-SIGN demonstrated with surface force measurements. *Proc. Natl. Acad. Sci. U S A,* (2009), **106**, 11524-11529.
176. W. I. Weis and K. Drickamer. Trimeric structure of a C-type mannose-binding protein. *Structure,* (1994), **2**, 1227-1240.
177. G. Tabarani, M. Thépaut, D. Stroebel, C. Ebel, C. Vivès, P. Vachette, D. Durand and F. Fieschi. DC-SIGN neck domain is a pH-sensor controlling oligomerization: SAXS and hydrodynamic studies of extracellular domain. *J. Biol. Chem.,* (2009), **284**, 21229-21240.
178. I. S. Trowbridge, J. F. Collawn and C. R. Hopkins. Signal-dependent membrane protein trafficking in the endocytic pathway. *Annu. Rev. Cell. BioI.,* (1993), **9**, 129-161.
179. H. Feinberg, Y. Guo, D. A. Mitchell, K. Drickamer and W. I. Weis. Extended neck regions stabilize tetramers of the receptors DC-SIGN and DC-SIGNR. *J. Biol. Chem.,* (2005), **280**, 1327-1335.
180. B. I. de Bakker, F. de Lange, A. Cambi, J. P. Korterik , E. M. van Dijk, N. F. van Hulst, C. G. Figdor and M. F. Garcia-Parajo. Nanoscale organization of the pathogen receptor DC-SIGN mapped by single-molecule high-resolution fluorescence microscopy. *Chemphyschem.,* (2007), **8**, 1473-1480.
181. C. Manzo, J. A. Torreno-Pina, B. Joosten, I. Reinieren-Beeren, E. J. Gualda, P. Loza-Alvarez, C. G. Figdor, M. F. Garcia-Parajo and A. Cambi. The neck region of the C-type lectin DC-SIGN regulates its surface spatiotemporal organization and virus-binding capacity on antigen-presenting cells. *J. Biol. Chem.,* (2012), **287**, 38946-38955.
182. A. Canales, I. Boos, L. Perkams, L. Karst, T. Luber, T. Karagiannis, G. Domínguez, F. J. Cañada, J. Pérez-Castells, D. Häussinger, C. Unverzagt and J. Jiménez-Barbero. Breaking the limits in analyzing carbohydrate recognition by NMR spectroscopy: Resolving branch-selective interaction of a tetra-antennary N-glycan with lectins. *Angew. Chem. Int. Ed. Engl.,* (2017), **56**, 14987-14991.
183. B. F. de Toro, W. Peng, A. J. Thompson, G. Domínguez, F. J. Cañada, J. Pérez-Castells, J. C. Paulson, J. Jiménez-Barbero and A. Canales. Avenues to Characterize the Interactions of Extended N-Glycans with Proteins by NMR Spectroscopy: The Influenza Hemagglutinin Case. *Angew. Chem. Int. Ed. Engl.,* (2018), **57**, 15051-15055.
184. A. Gimeno, S. Delgado, P. Valverde, S. Bertuzzi, M. A. Berbís, J. Echavarren, A. Lacetera, S. Martín-Santamaría, A. Surolia, F. J. Cañada, J. Jiménez-Barbero and A. Ardá. Minimizing the entropy penalty for ligand binding: lessons from the molecular recognition of the histo blood-group antigens by human galectin-3. *Angew. Chem. Int. Ed. Engl.,* (2019), **58**, 7268-7272.
185. J.-P. Gourdine, G. Cioci, L. Miguet, C. Unverzagt, D. V. Silva, A. Varrot, C. Gautier, E. J. Smith-Ravin and A. Imberty. High affinity interaction between a

bivalve C-type lectin and a biantennary complex-type n-glycan revealed by crystallography and microcalorimetry. *J. Biol. Chem.,* (2008), **283**, 30112-30120.

186. T. K. Dam, T. A. Gerken and C. F. Brewer. Thermodynamics of multivalent carbohydrate− lectin cross-linking interactions: Importance of entropy in the bind and jump mechanism. *Biochemistry,* (2009), **48**, 3822-3827.
187. A. Imberty, E. P. Mitchell and M. Wimmerová. Structural basis of high-affinity glycan recognition by bacterial and fungal lectins. *Curr. Opin. Struct. Biol*., (2005), **15**, 525-534.
188. S. Cecioni, A. Imberty and S. Vidal. Glycomimetics versus multivalent glycoconjugates for the design of high affinity lectin ligands. *Chem. Rev.,* (2015), **115**, 525-561.
189. J. Li, F. Yu, Y. Chen and D. Oupický. Polymeric drugs: Advances in the development of pharmacologically active polymers. *J. Control Release,* (2015), **219**, 369-382.
190. T. K. Dam and C. F. Brewer. Effects of clustered epitopes in multivalent ligand−receptor interactions. *Biochemistry,* (2008), **47**, 8470-8476.
191. J. J. Lundquist and J. E. Toone. The cluster glycoside effect. *Chem. Rev.,* (2002), **102**, 555-578.
192. V. Wittmann and R. J. Pieters. Bridging lectin binding sites by multivalent carbohydrates. *Chem. Soc. Rev.,* (2013), **42**, 4492-4503.
193. L. Herfurth, B. Ernst, B. Wagner, D. Ricklin, D. S. Strasser, J. L. Magnani, A. J. Benie and T. Peters. Comparative epitope mapping with Saturation Transfer Difference NMR of sialyl LewisA compounds and derivatives bound to a monoclonal antibody. *J. Med. Chem.,* (2005), **48**, 6879-6886.
194. E. J. Soilleux, R. Barten and J. Trowsdale. Cutting edge: DC-SIGN; a related gene, DC-SIGNR; and CD23 form a cluster on 19p13. *J. Immunol.,* (2000), **165**, 2937-2942.
195. E. van Liempt, A. Imberty, C. M. C. Bank, S. J. van Vliet, Y. van Kooyk, T. B. Geijtenbeek and I. van Die. Molecular basis of the differences in binding properties of the highly related C-type lectins DC-SIGN and L-SIGN to Lewis X trisaccharide and *Schistosoma mansoni* egg antigens. *J. Biol. Chem.,* (2004), **279**, 33161-33167.
196. Y. Guo, H. Feinberg, E. Conroy, D. A. Mitchell, R. Alvarez, O. Blixt, M. E. Taylor, W. I. Weis and K. Drickamer. Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat. Struct. Mol. Biol.,* (2004), **11**, 591–598.
197. H. Feinberg, A. S. Powlesland, M. E. Taylor and W. I. Weis. Trimeric structure of langerin. *J. Biol. Chem.,* (2010), **285**, 13285-13293.
198. L. Chatwell, A. Holla, B. B. Kaufer and A. Skerra. The carbohydrate recognition domain of Langerin reveals high structural similarity with the one of DC-SIGN but an additional, calcium-independent sugar-binding site. *Mol. Immunol.,* (2008), **45**, 1981-1994.

199. M. A. de Jong, L. E. Vriend, B. Theelen, M. E. Taylor, D. Fluitsma, T. Boekhout and T. B. Geijtenbeek. C-type lectin Langerin is a beta-glucan receptor on human Langerhans cells that recognizes opportunistic and pathogenic fungi. *Mol. Immunol.,* (2010), **47**, 1216-1225.
200. H. Feinberg, M. E. Taylor, N. Razi, R. McBride, Y. A. Knirel, S. A. Graham, K. Drickamer and W. I. Weis. Structural basis for langerin recognition of diverse pathogen and mammalian glycans through a single binding site. *J. Mol. Biol.,* (2011), **405**, 1027-1039.
201. J. C. Muñoz-García, E. Chabrol, R. R. Vivès, A. Thomas, J. L. de Paz, J. Rojo, A. Imberty, F. Fieschi, P. M. Nieto and J. Angulo. Langerin−Heparin interaction: two binding sites for small and large ligands as revealed by a combination of NMR spectroscopy and cross-linking mapping experiments. *J. Am. Chem. Soc.,* (2015), **137**, 4100-4110.
202. F. Ota, T. Hirayama, Y. Kizuka, Y. Yamaguchi, R. Fujinawa, M. Nagata, H. S. Ismanto, B. Lepenies, J. Aretz, C. Rademacher, P. H. Seeberger, T. Angata, S. Kitazume, K. Yoshida, T. Betsuyaku, K. Kida, S. Yamasaki and N. Taniguchi. High affinity sugar ligands of C-type lectin receptor Langerin. *Biochim. Biophys. Acta Gen. Subj.,* (2018), **1862**, 1592-1601.
203. J. Zhao, X. Liu, C. Kao, E. Zhang, Q. Li, F. Zhang and L. J. Linhardt. Kinetic and structural studies of interactions between glycosaminoglycans and Langerin. *Biochemistry,* (2016), **55**, 4552-4559.
204. W. I. Weis, K. Drickamer and W. A. Hendrickson. Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature,* (1992), **360**, 127-134.
205. K. K. Ng, A. R. Kolatkar, S. Park-Snyder, H. Feinberg, D. A. Clark, K. Drickamer and W. I. Weis. Orientation of bound ligands in mannose-binding proteins. Implications for multivalent ligand recognition. *J. Biol. Chem.,* (2002), **277**, 16088-16095.
206. E. W. Adams, D. M. Ratner, H. R. Bokesch, J. B. McMahon, B. R. O'Keefe and P. H. Seeberger. Oligosaccharide and glycoprotein microarrays as tools in HIV glycobiology: glycan-dependent gp120/protein interactions. *Chem. Biol.,* (2004), **11**, 875–881.
207. H. Feinberg, D. A. Mitchell, K. Drickamer and W. I. Weis. Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science,* (2001), **294**, 2163–2166.
208. J. J. Reina, I. Diaz, P. M. Nieto, N. E. Campillo, J. A. Paez, G. Tabarani, F. Fieschi and J. Rojo. Docking, synthesis, and NMR studies of mannosyl trisaccharide ligands for DC-SIGN lectin. *Org. Biomol. Chem.,* (2008), **6**, 2743–2754.
209. A. R. Temming, G. Dekkers, F. S. van de Bovenkamp, H. R. Plomp, A. E. H. Bentlage, Z. Szittner, N. I. L. Derksen, M. Wuhrer, T. Rispens and G. Vidarsson. Human DC-SIGN and CD23 do not interact with human IgG. *Sci. Rep.,* (2019), **9**, 9995.

210. K. Brzezicka, B. Echeverria, S. Serna, A. van Diepen, C. H. Hokke and N.-C. Reichardt. Synthesis and microarray-assisted binding studies of core xylose and fucose containing N-glycans. *ACS Chem. Biol.,* (2015), **10**, 1290–1302.

211. B. Echeverria, S. Serna, S. Achilli, C. Vivès, J. Pham, M. Thépaut, C. H. Hokke, F. Fieschi and N.-C. Reichardt. Chemoenzymatic synthesis of N‑glycan positional isomers and evidence for branch selective binding by monoclonal antibodies and human C‑type lectin receptors. *ACS Chem. Biol.,* (2018), **13**, 2269−2279.

212. J. Angulo, I. Diaz, J. J. Reina, G. Tabarani, F. Fieschi, J. Rojo and P. M. Nieto. Saturation Transfer Difference (STD) NMR spectroscopy characterization of dual binding mode of a mannose disaccharide to DC-SIGN *ChemBioChem,* (2008), **9**, 2225–2227.

213. H. Feinberg, R. Castelli, K. Drickamer, P. H. Seeberger and W. I. Weis. Multiple modes of binding enhance the affinity of DC-SIGN for high mannose N-linked glycans found on viral glycoproteins. *J. Biol. Chem.,* (2007), **282**, 4202–4209.

214. E. van Liempt, C. M. C. Bank, P. Mehta, J. J. García-Vallejo, Z. S. Kawar, R. Geyer, R. A. Alvarez, R. D. Cummings, Y. van Kooyk and I. van Die. Specificity of DC-SIGN for mannose- and fucose-containing glycans. *FEBS Lett.,* (2006), **580**, 6123–6131.

215. A. Holla and A. Skerra. Comparative analysis reveals selective recognition of glycans by the dendritic cell receptors DC-SIGN and Langerin. *Protein Eng. Des. Sel.,* (2011), **24**, 659-669.

216. S. Shahzad-ul-Hussan, M. Sastry, T. Lemmin, C. Soto, S. Löesgen, D. A. Scott, J. R. Davison, R. O'Connor, Pe. D. Kwong and C. A. Bewley. NMR insights into the conformational properties of Man-9 and its recognition by two HIV binding proteins. *Chembiochem.,* (2017), **18**, 764-771.

217. S. Pitarque, J.-L. Herrmann, J.-L. Duteyrat, M. Jackson, G. R. Stewart, F. Lecointe, B. Payre, O. Schwartz, D. B. Young, G. Marchal, P. H. Lagrange, G. Puzo, B. Gicquel, J. Nigou and O. Neyrolles. Deciphering the molecular bases of *Mycobacterium tuberculosis* binding to the lectin DC-SIGN reveals an underestimated complexity. *J. Biochem.,* (2005), **392**, 615–624.

218. J. te Riet, I. Reinieren-Beeren, C. G. Figdor and A. Cambi. AFM force spectroscopy reveals how subtle structural differences affect the interaction strength between *Candida albicans* and DC-SIGN. *J. Mol. Recognit.,* (2015), **28**, 687–698.

219. T. van Montfort, D. Eggink, M. Boot, M. Tuen, C. E. Hioe, B. Berkhout and R. W. Sanders. HIV-1 N-glycan composition governs a balance between dendritic cell-mediated viral transmission and antigen presentation. *J. Immunol.,* (2011), **187**, 4676-4685.

220. F. Probert, S. B.-M. Whittaker, M. Crispin, D. A. Mitchell and A. M. Dixon. Solution NMR analyses of the C-type carbohydrate recognition domain of DC-SIGNR protein reveal different binding modes for HIV-derived oligosaccharides and smaller glycan fragments. *J. Biol. Chem.,* (2013), **288**, 22745-22757.

221. B. J. Appelmelk, I. van Die, S. J. van Vliet, C. M. J. E. Vandenbroucke-Grauls, T. B. Geijtenbeek and Y. van Kooyk. Cutting edge: carbohydrate profiling identifies new pathogens that interact with dendritic cell-specific ICAM-3-grabbing nonintegrin on dendritic cells. *J. Immunol.,* (2003), **170**, 1635−1639.
222. K. Pederson, D. A. Mitchell, and J. H. Prestegard. Structural characterization of the DC-SIGN−LewisX complex. *Biochemistry,* (2014), **53**, 5700−5709.
223. W. S. Somers, J. Tang, G. D. Shaw and R. T. Camphausen. Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1. *Cell,* (2000), **103**, 467–479.
224. H. F. Azurmendi, M. Martin-Pastor and C. A. Bush. Conformational studies of Lewis X and Lewis A trisaccharides using NMR residual dipolar couplings. *Biopolymers,* (2002), **63**, 89−98.
225. M. D. Battistel, H. F. Azurmendi, M. Frank and D. I. Freedberg. Uncovering nonconventional and conventional hydrogen bonds in Oligosaccharides through NMR experiments and molecular modeling: application to Sialyl Lewis-X. *J. Am. Chem. Soc.,* (2015), **137**, 13444−13447.
226. a) A. Varki. Biological roles of glycans. *Glycobiology,* (2017), **27**, 3-49. b) M. E. Taylor and K. Drickamer. Paradigms for glycan-binding receptors in cell adhesion. *Curr. Opin. Cell. Biol.,* (2007), **19**, 572-577.
227. a) K. Scheffler. B. Ernst, A. Katopodis, J. L. Magnani, W. Wang, R. Weiseinann and T. Peters. Determination of the bioactive conformation of the carbohydrate ligand in the E-selectin/sialyl LewisY complex. *Angew. Chem. Int. Ed. Engl.,* (1995), **34**, 1841-1844.
228. K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.,* (2008), **29**, 622-655.
229. S. Meyer, E. van Liempt, A. Imberty, Y. van Kooyk, H. Geyer, R. Geyer and I. van Die. DC-SIGN mediates binding of dendritic cells to authentic pseudo-LewisY glycolipids of *Schistosoma mansoni Cercariae*, the first parasite-specific ligand of DC-SIGN. *J. Biol. Chem.,* (2005), **280**, 37349-37359.
230. E. Rodríguez, S. T. T. Schetters and Y. van Kooyk. The tumour glyco-code as a novel immune checkpoint for immunotherapy. *Nat. Rev. Immunol.,* (2018), **18**, 204-211.
231. A. J. Noll, Y. Yu, Y. Lasanajak, G. Duska-McEwen, R. H. Buck, D. F. Smith and R. D. Cummings. Human DC-SIGN binds specific human milk glycans. *Biochem. J.,* (2016), **473**, 1343–1353.
232. M. A. Naarding, I. S. Ludwig, F. Groot, B. Berkhout, T. B. Geijtenbeek, G. Pollakis and W. A. Paxton. Lewis X component in human milk binds DC-SIGN and inhibits HIV-1 transfer to CD4+ T lymphocytes. *J. Clin. Invest.,* (2005), **115**, 3256-3264.
233. N. Koning, S. F. M. Kessen, J. P. van der Voorn, B. J. Appelmelk, P. V. Jeurink, L. M. J. Knippels, J. Garssen and Y. van Kooyk. Human milk blocks DC-SIGN–pathogen interaction via MUC1. *Front. Immunol.,* (2015), **6**, 112.

234. D. Serrano-Gómez, R. T. Martínez-Nuñez, E. Sierra-Filardi, N. Izquierdo, M. Colmenares, J. Pla, L. Rivas, J. Martinez-Picado, J. Jimenez-Barbero, J. Alonso-Lebrero, S. González and A. L. Corbí. AM3 modulates dendritic cell pathogen recognition capabilities by targeting DC-SIGN. *Antimicrob. Agents Chemother.,* (2007), **51**, 2313-2323.

235. J. M. Silva, E. Zupancic, G. Vandermeulen, V. G. Oliveira, A. Salgado, M. Videira, M. Gaspar, L. Graca, V. Préat and H. F. Florindo. In vivo delivery of peptides and Toll-like receptor ligands by mannose-functionalized polymeric nanoparticles induces prophylactic and therapeutic anti-tumor immune responses in a melanoma model. *J. Control Release,* (2015), **198**, 91-103.

236. M. A. Boks, M. Ambrosini, S. C. Bruijns, H. Kalay, L. van Bloois, G. Storm, J. J. Garcia-Vallejo and Y. van Kooyk. MPLA incorporation into DC-targeting glycoliposomes favours anti-tumour T cell responses. *J. Control Release,* (2015), **216**, 37-46.

237. B. M. Illescas, J. Rojo, R. Delgado and N. Martín. Multivalent glycosylated nanostructures to inhibit Ebola virus infection. *J. Am. Chem. Soc*., (2017), **139**, 6018−6025.

238. D. A. Mitchell, N. A. Jones, S. J. Hunter, J. Cook, S. F. Jenkinson, M. R. Wormald, R. A. Dwek and G. W. J. Fleet. Synthesis of 2-C-branched derivatives of D-mannose: 2-C-aminomethyl-D-mannose binds to the human C-type lectin DC-SIGN with affinity greater than an order of magnitude compared to that of D-mannose. *Tetrahedron Asymmetry,* (2007), **18**, 1502−1510.

239. S. Mari, H. Posteri, G. Marcou, D. Potenza, F. Micheli, F. Javier Cañada, J. Jimenez-Barbero and A. Bernardi. Synthesis, conformational studies and mannosidase stability of a mimic of 1,2-mannobioside. *Eur. J. Org. Chem.,* (2004), 5119-5125.

240. V. Bordoni, V. Porkolab, S. Sattin, M. Thépaut, I. Frau, L. Favero, P. Crotti, A. Bernardi, F. Fieschi and V. Di Bussolo. Stereoselective innovative synthesis and biological evaluation of new real carba analogues of minimal epitope Mana(1,2)Man as DC-SIGN inhibitors. *RSC Adv.,* (2016), **6**, 89578-89584.

241. J. J. Reina, S. Sattin, D. Invernizzi, S. Mari, L. Martinez-Prats, G. Tabarani, F. Fieschi, R. Delgado, P. M. Nieto, J. Rojo and A. Bernardi. 1,2-mannobioside mimic: synthesis, DC-SIGN interaction by NMR and docking, and antiviral activity. *ChemMedChem,* (2007), **2**, 1030–1036.

242. M. Thépaut, C. Guzzi, I. Sutkeviciute, S. Sattin, R. Ribeiro-Viana, N. Varga, E. Chabrol, J. Rojo, A. Bernardi, J. Angulo, P. M. Nieto and F. Fieschi, F. Structure of a glycomimetic ligand in the carbohydrate recognition domain of C-type lectin DC-SIGN. Structural requirements for selectivity and ligand design. *J. Am. Chem. Soc.,* (2013), **135**, 2518−2529.

243. N. Varga, I. Sutkeviciute, C. Guzzi, J. McGeagh, I. Petit-Haertlein, S. Gugliotta, J. Weiser, J. Angulo, F. Fieschi and A. Bernardi. Selective targeting of dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin (DC-SIGN) with mannose-based glycomimetics: synthesis and interaction studies of

bis(benzylamide) derivatives of a pseudomannobioside. *Chem. Eur. J.,* (2013), **19**, 4786–4797.

244. N. Obermajer, S. Sattin, C. Colombo, M. Bruno, U. Svajger, M. Anderluh and A. Bernardi. Design, synthesis and activity evaluation of mannose-based DC-SIGN antagonists. *Mol. Diversity,* (2011), **15**, 347–360.
245. V. Porkolab, E. Chabrol, N. Varga, S. Ordanini, I. Sutkevičiūtė, M. Thépaut, M. J. García-Jiménez, E. Girard, P. M. Nieto, A. Bernardi and F. Fieschi. Rational-differential design of highly specific glycomimetic ligands: targeting DC-SIGN and excluding langerin recognition. *ACS Chem. Biol.*, (2018), **13**, 600−608.
246. L. Medve, S. Achilli, J. Guzman-Caldentey, M. Thépaut, L. Senaldi, A. Le Roy, S. Sattin, C. Ebel, C. Vivès, S. Martin-Santamaria, A. Bernardi and F. Fieschi. Enhancing potency and selectivity of a DC-SIGN glycomimetic ligand by fragment-based design: structural basis. *Chem. Eur. J.,* (2019), **25**, 14659 - 14668.
247. C. Guzzi, J. C. Muñoz-García, P. M. Enriquez-Navas, J. Rojo, J. Angulo and P. M. Nieto. NMR studies on carbohydrate interactions with DC-SIGN towards a quantitative STD analysis. *Pure Appl. Chem.,* (2013), **85**, 1771–1787.
248. C. Guzzi, P. Alfarano, I. Sutkeviciute, S. Sattin, R. Ribeiro-Viana, F. Fieschi, A. Bernardi, J. Weiser, J. Rojo, J. Angulo and P. M. Nieto. Detection and quantitative analysis of two independent binding modes of a small ligand responsible for DC-SIGN clustering. *Org. Biomol. Chem.,* (2016), **14**, 335-344.
249. I. Sutkeviciute, M. Thépaut, S. Sattin, A. Berzi, J. McGeagh, S. Grudinin, J. Weiser, A. Le Roy, J. J. Reina, J. Rojo, M. Clerici, A. Bernardi, C. Ebel and F. Fieschi. Unique DC-SIGN clustering activity of a small glycomimetic: a lesson for ligand design. *ACS Chem. Biol.,* (2014), **9**, 1377–1385.
250. S. Sattin, A. Daghetti, M. Thepaut, A. Berzi, M. Sanchez-Navarro, G. Tabarani, J. Rojo, F. Fieschi, M. Clerici and A. Bernardi. Inhibition of DC-SIGN-mediated HIV infection by a linear trimannoside mimic in a tetravalent presentation. *ACS Chem. Biol.,* (2010), **5**, 301–312.
251. A. Berzi, J. J. Reina, R. Ottria, I. Sutkeviciute, P. Antonazzo, M. Sanchez-Navarro, E. Chabrol, M. Biasin, D. Trabattoni, I. Cetin, J. Rojo, F. Fieschi, A. Bernardi and M. Clerici. A glycomimetic compound inhibits DC-SIGN-mediated HIV infection in cellular and cervical explant models. *AIDS*, (2012), **26**, 127–137.
252. N. Varga, I. Sutkeviciute, R. Ribeiro-Viana, A. Berzi, R. Ramdasi, A. Daghetti, G. Vettoretti, A. Amara, M. Clerici, J. Rojo, F. Fieschi and A. Bernardi. A multivalent inhibitor of the DC-SIGN dependent uptake of HIV-1 and Dengue virus. *Biomaterials*, (2014), **35**, 4175−4184.
253. J. Luczkowiak, S. Sattin, I. Sutkeviciute, J. Juan Reina, M. Sanchez-Navarro, M. Thepaut, L. Martinez-Prats, A. Daghetti, F. Fieschi, R. Delgado, A. Bernardi and J. Rojo. Pseudosaccharide functionalized dendrimers as potent inhibitors of DC-SIGN dependent Ebola pseudotyped viral infection. *Bioconjugate Chem.,* (2011), **22**, 1354–1365.

254. S. Ordanini, N. Varga, V. Porkolab, M. Thépaut, L. Belvisi, A. Bertaglia, A. Palmioli, A. Berzi, D. Trabattoni, M. Clerici, F. Fieschi and A. Bernardi. Designing nanomolar antagonists of DC-SIGN-mediated HIV infection: ligand presentation using molecular rods. *Chem. Commun.,* (2015), **51**, 3816–3819.

255. A. Berzi, S. Ordanini, B. Joosten, D. Trabattoni, A. Cambi, A. Bernardi and M. Clerici. Pseudo-mannosylated DC-SIGN ligands as immunomodulants. *Sci. Rep.,* (2016), **6**, 35373.

256. T. Tomasic, D. Hajsek, U. Svajger, J. Luzar, N. Obermajer, I. Petit-Haertlein, F. Fieschi and M. Anderluh. Monovalent mannose-based DC-SIGN antagonists: Targeting the hydrophobic groove of the receptor. *Eur. J. Med. Chem.,* (2014), **75**, 308–326.

257. A. Kotar, T. Tomašič, M. L. Živkovid, G. Jug, J. Plavec, and M. Anderluh. STD NMR and molecular modelling insights into interaction of novel mannose-based ligands with DC-SIGN. *Org. Biomol. Chem.,* (2016), **14**, 862-875.

258. G. Timpano, G. Tabarani, M. Anderluh, D. Invernizzi, F. Vasile, D. Potenza, P. M. Nieto, J. Rojo, F. Fieschi and A. Bernardi. Synthesis of novel DC-SIGN ligands with an α-fucosylamide anchor. *ChemBioChem,* (2008), **9**, 1921–1930.

259. M. Andreini, D. Doknic, I. Sutkeviciute, J. J. Reina, J. Duan, E. Chabrol, M. Thepaut, E. Moroni, F. Doro, L. Belvisi, J. Weiser, J. Rojo, F. Fieschi and A. Bernardi. Second generation of fucose-based DC-SIGN ligands: affinity improvement and specificity versus Langerin. *Org. Biomol. Chem.,* (2011), **9**, 5778-5786.

260. D. Doknic, M. Abramo, I. Sutkeviciute, A. Reinhardt, C. Guzzi, M. K. Schlegel, D. Potenza, P. M. Nieto, F. Fieschi, P. H. Seeberger and A. Bernardi. Synthesis and characterization of Linker-armed fucose-based glycomimetics. *Eur. J. Org. Chem.,* (2013), 5303–5314.

261. C. Guzzi, J. Angulo, F. Doro, J. J. Reina, M. Thepaut, F. Fieschi, A. Bernardi, J. Rojo and P. M. Nieto. Insights into molecular recognition of Lewis X mimics by DC-SIGN using NMR and molecular modeling. *Org. Biomol. Chem.,* (2011), **9**, 7705–7712.

262. H. Lis and N. Sharon. Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.,* (1998), **98**, 637-674.

263. E. R. Li, T. P. Hogervorst, S. Achilli, S. C. Bruijns, T. Arnoldus, C. Vivès, C. C. Wong, M. Thépaut, N. J. Meeuwenoord, H. Elst van den, H. S. Overkleeft, G. A. Marel van der, D. V. Filippov, S. J. van Vliet, F. Fieschi, J. D. Codée and Y. Van Kooyk. Systematic dual targeting of dendritic cell C-type lectin receptor DC-SIGN and TLR7 using a trifunctional mannosylated antigen. *Front. Chem.,* (2019), **7**, 650.

264. F. Lasala, E. Arce, J. R. Otero, J. Rojo and R. Delgado. Mannosyl glycodendritic structure inhibits DC-SIGN-mediated Ebola virus infection in *cis* and in *trans*. *Antimicrob. Agents Chemother.,* (2003), **47**, 3970-3972.

265. J. Rojo and R. Delgado. Glycodendritic structures: promising new antiviral drugs. *J. Antimicrob. Chemother.,* (2004), **54**, 579–581.

266. G. Tabarani, J. J. Reina, C. Ebel, C. Vivès, H. Lortat-Jacob, J. Rojo and F. Fieschi. Mannose hyperbranched dendritic polymers interact with clustered organization of DC-SIGN and inhibit gp120 binding. *FEBS Lett.,* (2006), **580**, 2402–2408.
267. E. Blattes, A. Vercellone, H. Eutamène, C.-O. Turrin, V. Théodorou, J.-P. Majorai, A.-M. Caminade, J. Prandi, J. Nigou and G. Puzo. Mannodendrimers prevent acute lung inflammation by inhibiting neutrophil recruitment. *Proc. Natl. Acad. Sci.,* (2013), **110**, 8795-8800.
268. S.-K.Wang, P.-H. Liang, R. D. Astronomo, T.-L. Hsu, S.-L. Hsieh, D. R. Burton and C.-H. Wong. Targeting the carbohydrates on HIV-1: Interaction of oligomannose dendrons with human monoclonal antibody 2G12 and DC-SIGN. *Proc. Natl. Acad. Sci. U S A,* (2008), **105**, 3690-3695.
269. O. Martínez-Ávila, L. M. Bedoya, M. Marradi, C. Clavel, J. Alcamí and S. Penadés. Multivalent manno-glyconanoparticles inhibit DC-SIGN-mediated HIV-1 trans-infection of human T cells. *ChemBioChem.,* (2009), **10**, 1806-1809.
270. O. Martinez-Avila, K. Hijazi, M. Marradi, C. Clavel, C. Campion, C. Kelly and S. Penades. Gold manno-glyconanoparticles: multivalent systems to block HIV-1 gp120 binding to the lectin DC-SIGN. *Chem. Eur. J.,* (2009), **15**, 9874–9888.
271. S. Brument, C. Cheneau, Y. Brissonnet, D. Deniaud, F. Halary and S. G. Gouin. Polymeric mannosides prevent DC-SIGN-mediated cell-infection by cytomegalovirus. *Org. Biomol. Chem.,* (2017), **15**, 7660–7671.
272. J. Luczkowiak, A. Muñoz, M. Sánchez-Navarro, R. Ribeiro-Viana, A. Ginieis, B. M. Illescas, N. Martín, R. Delgado and J. Rojo. Glycofullerenes inhibit viral infection*. Biomacromolecules,* (2013), **14**, 431-437.
273. A. Muñoz, D. Sigwalt, B. M. Illescas, J. Luczkowiak, L. Rodríguez-Pérez, I. Nierengarten, M. Holler, J.-S. Remy, K. Buffet, S. P. Vincent, J. Rojo, R. Delgado, J.-F. Nierengarten and N. Martín. Synthesis of giant globular multivalent glycofullerenes as potent inhibitors in a model of Ebola virus infection. *Nat. Chem.,* (2016), **8**, 50-57.
274. J. J. Reina, O. S. Maldonado, G. Tabarani, F. Fieschi and J. Rojo. Mannose glycoconjugates functionalized at positions 1 and 6. Binding analysis to DC-SIGN using biosensors. *Bioconjugate Chem.,* (2007), **18**, 963–969.
275. B. Bertolotti, I. Sutkeviciute, M. Ambrosini, R. Ribeiro-Viana, J. Rojo, F. Fieschi, H. Dvořáková, M. Kašáková, K. Parkan, M. Hlaváčková, K. Nováková and J. Moravcová. Polyvalent C-glycomimetics based on L-fucose or D-mannose as potent DC-SIGN antagonists. *Org. Biomol. Chem.,* (2017), **15**, 3995-4004.
276. B. Bertolotti, B. Oroszová, I. Sutkeviciute, L. Kniezo, F. Fieschi, K. Parkan, Z. Lovyová, M. Kašáková and J. Moravcová. Nonhydrolyzable C-disaccharides, a new class of DC-SIGN ligands. *Carbohydr. Res.,* (2016), **435**, 7-18.
277. I. Morbioli, V. Porkolab, A. Magini, A. Casnati, F. Fieschi and F. Sansone. Mannosylcalix[n]arenes as multivalent ligands for DC-SIGN. *Carbohydr. Res.,* (2017), **453-454**, 36-43.

278. L. Dehuyser, E. Schaeffer, O. Chaloin, C. G. Mueller, R. Baati and A. Wagner. Synthesis of novel mannoside glycolipid conjugates for inhibition of HIV-1 trans-infection. *Bioconjug. Chem.,* (2012), **23**, 1731-1739.

279. N. Frison, M. E. Taylor, E. Soilleux, M. T. Bousser, R. Mayer, M. Monsigny, K. Drickamer and A. C. Roche. Oligolysine-based oligosaccharide clusters: selective recognition and endocytosis by the mannose receptor and dendritic cell-specific intercellular adhesion molecule 3 (ICAM-3)-grabbing nonintegrin. *J. Biol. Chem.,* (2003), **278**, 23922-23929.

280. R. Ribeiro-Viana, M. Sánchez-Navarro, J. Luczkowiak, J. R. Koeppe, R. Delgado, J. Rojo and B. G. Davis. Virus-like glycodendrinanoparticles displaying quasi-equivalent nested polyvalency upon glycoprotein platforms potently block viral infection. *Nat. Commun.,* (2012), **3**, 1303.

281. D. A. Mitchell, Q. Zhang, L. Voorhaar, D. M. Haddleton, S. Herath, A. S. Gleinich, H. S. Randeva, M. Crispin, H. Lehnert, R. Wallis, S. Patterson and C. R. Becer. Manipulation of cytokine secretion in human dendritic cells using glycopolymers with picomolar affinity for DC-SIGN. *Chem. Sci.,* (2017), **8**, 6974-6980.

282. M. Ciobanu, K.-T. Huang, J.-P. Daguer, S. Barluenga, O. Chaloin, E. Schaeffer, C. G. Mueller, D. A. Mitchell and N. Winssinger. Selection of a synthetic glycan oligomer from a library of DNA-templated fragments against DC-SIGN and inhibition of HIV gp120 binding to dendritic cells. *Chem. Commun.,* (2011), **47**, 9321–9323.

283. K. Gorska, K. T. Huang, O. Chaloin and N. Winssinger. DNA-templated homo- and heterodimerization of peptide nucleic acid encoded oligosaccharides that mimick the carbohydrate epitope of HIV. *Angew. Chem. Int. Ed.,* (2009), **48**, 7695-7700.

284. J. Huang, Q. Zhang, G.-Z. Li, D. M. Haddleton, R. Wallis, D. Mitchell, A. Heise and C. R. Becer. Synthetic glycopolypeptides as potential inhibitory agents for dendritic cells and HIV-1 trafficking. *Macromol. Rapid Commun.,* (2013), **34**, 1542−1546.

285. C. R. Becer, M. I. Gibson, J. Geng, R. Ilyas, R. Wallis, D. A. Mitchell and D. M. Haddleton. High-affinity glycopolymer binding to human DC-SIGN and disruption of DC-SIGN interactions with HIV envelope glycoprotein. *J. Am. Chem. Soc.,* (2010), **132**, 15130-15132.

286. Q. Zhang, J. Collins, A. Anastasaki, R. Wallis, D. A. Mitchell, C. R. Becer and D. M. Haddleton. Sequence-controlled multi-block glycopolymers to inhibit DC-SIGN-gp120 binding. *Angew. Chem. Int. Ed.*, (2013), **52**, 4435-4439.

287. Q. Zhang, L. Su, J. Collins, G. Chen, R. Wallis, D. A. Mitchell, D. M. Haddleton and C. R. Becer. Dendritic cell lectin-targeting sentinel-like unimolecular glycoconjugates to release an anti-HIV drug. *J. Am. Chem. Soc.,* (2014), **136**, 4325−4332.

288. A. Le Moignic, V. Malard, T. Benvegnu, L. Lemiègre, M. Berchel, P.-A. Jaffrès, C. Baillou, M. Delost, R. Macedo, J. Rochefort, G. Lescaille, C. Pichon, F. M.

Lemoine, P. Midoux and V. Mateo. Preclinical evaluation of mRNA trimannosylated lipopolyplexes as therapeutic cancer vaccines targeting dendritic cells. *J. Control. Release*, (2018), **278**, 110-121.

289. W. W. J. Unger, A. J. van Beelen, S. C. Bruijns, M. Joshi, C. M. Fehres, L. van Bloois, M. I. Verstege, M. Ambrosini, H. Kalay, K. Nazmi, J. G. Bolscher, E. Hooijberg, T. D. de Gruijl, G. Storm and Y. van Kooyk. Glycan-modified liposomes boost CD4+ and CD8+ T-cell responses by targeting DC-SIGN on dendritic cells. *J. Control. Release,* (2012), **160,** 88–95.
290. C. M. Fehres, H. Kalay, S. C. M. Bruijns, S. A. M. Musaafir, M. Ambrosini, L. van Bloois, S. J. van Vliet, G. Storm, J. J. Garcia-Vallejo and Y. van Kooyk. Cross-presentation through Langerin and DC-SIGN targeting requires different formulations of glycan-modified antigens. *J. Control. Release,* (2015), **203**, 67–76.
291. K. C. Garber, K. Wangkanont, E. E. Carlson and L. L. Kiessling. A general glycomimetic strategy yields non-carbohydrate inhibitors of DC-SIGN. *Chem. Commun. (Camb).,* (2010), **46**, 6747-6749.
292. L. R. Prost, J. C. Grim, M. Tonelli and L. L. Kiessling. Noncarbohydrate glycomimetics and glycoprotein surrogates as DC-SIGN antagonists and agonists. *ACS Chem. Biol.,* (2012), **7**, 1603-1608.
293. M. J. Borrok and L. L. Kiessling. Non-carbohydrate inhibitors of the lectin DC-SIGN. *J. Am. Chem. Soc.,* (2007), **129**, 12780-12785.
294. S. L. Mangolg, L. R. Prost and L. L. Kiessling. Quinoxalinone Inhibitors of the Lectin DC-SIGN. *Chem. Sci.,* (2012), **3**, 772-777.
295. J. Aretz, E. C. Wamhoff, J. Hanske, D. Heymann and C. Rademacher. Computational and experimental prediction of human C-type lectin receptor druggability. *Front. Immunol.,* (2014), **5**, 323.
296. J. Aretz, Y. Kondoh, K. Honda, U. R. Anumala, M. Nazaré, N. Watanabe, H. Osada and C. Rademacher. Chemical fragment arrays for rapid druggability assessment. *Chem. Commun. (Camb).,* (2016), **52**, 9067-9070.
297. J. Schulze, H. Baukmann, R. Wawrzinek, F. F. Fuchsberger, E. Specker, J. Aretz , M. Nazaré and C. Rademacher. CellFy: A Cell-Based Fragment Screen against C-Type Lectins. *ACS Chem. Biol.,* (2018), **13**, 3229-3235.
298. J. Aretz, H. Baukmann, E. Shanina, J. Hanske, R. Wawrzinek, V. A. Zapolskii, P. H. Seeberger, D. E. Kaufmann and C. Rademacher. Identification of multiple druggable secondary sites by fragment screening against DC-SIGN. *Angew. Chem. Int. Ed. Engl.,* (2017), **56**, 7292-7296.

CHAPTER 2: OBJECTIVES

2. Objectives

The primordial scientific aim of this thesis project is the characterization of the fine details of the interaction processes between DC-SIGN and carbohydrate antigens, especially those bearing Fucose units. From a biological perspective, these data will provide a better understanding of cancer, allergy and other autoimmune diseases in which DC-SIGN takes part through the recognition of endogenous fucosylated substrates.

This scientific objective is intimately related to the training aims, focused on the acquisition of high-level knowledge in chemical biology, including synthesis, conformational analysis, and molecular recognition, especially focused in state-of-the-art NMR methods and their applications to problems of biological interest as well as in their combination with computational tools.

Specifically, the first goal is to precisely describe the interaction of DC-SIGN with the histo blood group antigens A and B from a molecular point of view, thereby expanding the scope of the therapeutics development towards the rational design of Fuc-based mimetics.

In addition, the second important goal, now from the methodological perspective, concerns the optimization of a robust NMR approach that can provide the key features of protein/glycan interactions at atomic resolution, demonstrating its potential for its application in future studies with different sugar antigens and lectins.

As third and additional goal, the NMR protocol specifically optimized for DC-SIGN has been applied to the analysis of its interaction with two distinct fucosylated ligands, LDNF and LDN-DF, thus demonstrating the effectiveness of the NMR methodologies and providing insights into the interaction of DC-SIGN with *S. Mansoni* parasites.

CHAPTER 3: METHODS

Nowadays, the range of biophysical techniques and spectroscopic methodologies applicable to the study of protein-ligand interactions is tremendously wide. Many of these approaches directly rely on the observation of a measurable variable manifested by one entity (ligand or receptor) whose value is somewhat related to the presence (concentration) of its counterpart. This variable is usually derived from a spectroscopic property affected by the interaction. Thus, a great deal of the analytical biochemistry assays (lectin arrays [1], glycan arrays [2-3], ELISA [4]…) would be classified into this category, since they are generally based on measuring the absorption, transmittance, emission or dispersion of electromagnetic radiations with different frequencies (fluorescence emission [5], UV/Vis absorption, interference patterns [6], radioisotopes [7]…). Clearly, the configuration of an array is very versatile, as the measured parameter can be monitored directly arising from the studied system or indirectly from other sources (a third molecular entity behaving as competitor or probe, a subsequent chemical reaction providing active species, etc.). However, in most of cases, the scope of these techniques ends up at this point. They can be very potent tools to quickly screen protein-ligand interactions and infer the corresponding affinities. However, on the other side, no further information is obtained at atomic level, setting aside crucial structural factors that may explain the differences or similarities between the calculated affinities and better understand the global recognition event.

In the context of protein-carbohydrate interactions, Surface Plasmon Resonance (SPR) [8] or Isothermal Titration Calorimetry (ITC) [9] are common examples of methodologies employed, besides the aforementioned analytical assays. Again, they both entail the same limitations in the kind of information that they provide, although ITC stands out by offering additional kinetic and thermodynamic parameters to better describe the binding event. From a strict structural point of view, NMR and X-Ray are the most widely known methodologies to gather information from the molecular perspective [10-11]. Nowadays, cryo-Electron Microscopy should be added to this list, as it is an emerging tool in the analysis of protein-sugar interactions [12]. Undoubtedly, although X-Ray keeps being the reference technique for describing protein-ligand complexes with a high degree of detail (atomic resolution) [13], cryo-EM is also becoming a powerful tool [14], whereas structural determinations by NMR are mainly limited by its low intrinsic sensibility. Even though, NMR provides two crucial advantages that complement those drawbacks normally found in X-Ray studies [15], and enormously expand its applications for biological studies:

- The type of information: both X-Ray and cryo-EM approaches require the sample to be immobilized (low temp., vitrified) to generate the experimental dataset [16-18]. In terms of molecular dynamics, the protein-ligand complex is "frozen", what means that the built model exclusively represents a "frame" of the real system, ignoring molecular motions (rotations, torsions, bond vibrations). Otherwise, these techniques are blind to additional phenomena which only take place when considering a true dynamic system (cooperative effects, entropic effects from flexibility changes or enthalpic effects from dynamic interactions…). In this regard, solution NMR fully addresses the study of protein-ligand interactions from a dynamic perspective, getting closer to the real biological conditions.
- Secondly, the versatility of NMR provides several descriptive parameters (chemical shifts, couplings, NOE, relaxation…) that can be exploited in many ways depending on the studied nuclei and the configuration of the pulse sequences. In turn, such versatility has given rise to tens of experiments designed for distinct purposes [19]: characterization of unknown molecules (organic synthesis, natural products…), structure elucidation [20-21], ligand screening in drug discovery [22], estimation of affinities [23-24], conformational information [25-26], reaction kinetics [27], description of interactions at a molecular level [28-29]…

As mentioned above, X-Ray methodologies are frequently employed to quickly obtain a descriptive model of the system of interest, given that the dataset is directly translated into spatial coordinates of the heavy atoms (C, O, N, S...), whereas the interpretation of NMR data is not so straightforward but can likewise lead to the same kind of details. Consequently, the current practices often make use of both techniques at once, as they nicely complement each other [30].

In this work, NMR has been used as the reference technique to derive the essential data, assisted by theoretical calculations made by computational methods (Molecular Dynamics). The already existing X-Ray structures of DC-SIGN and related receptors have been suitably applied to provide the proper framework for the spatial interpretation and justification of the NMR data. Now, the NMR methods used along this thesis are briefly summarized in the next sections prior to the detailed analysis of the results. Also, a quick description of the physical basis behind NMR is also provided below.

3.1. NMR basics

Nuclear Magnetic Resonance (NMR) is a spectroscopic technique firstly described in 1938 by I. Rabi, although F. Bloch and E. M. Purcell were the first ones who applied it to the observation of liquids and solids (1946). For such a contribution, they were awarded with the Nobel Prize in Physics six years later (1952).

The physical phenomenon of NMR comes from an intrinsic property of the atomic nuclei: the ***nuclear spin*** [31-33]. Each spinning nucleus has an associated magnetic moment (μ), whose value is quantified by means of the nuclear spin quantum number (***I***) (Figure 3.1). This parameter takes integer values for those nuclei with an even mass number and half integer values for odd mass numbers. Moreover, when both mass and atomic numbers are even (^{12}C, ^{16}O), then I = 0 and these nuclei are consequently not observable by NMR. For other nuclei with non-zero I values, there are 2I+1 possible magnetic moment states, represented by the quantum number "m" ranging from –I to +I. Hence, the "z" component of each state is described by Eq. 1:

$$\mu_z = \frac{\gamma \, m \, h}{2\pi} \qquad (Eq. 1)$$

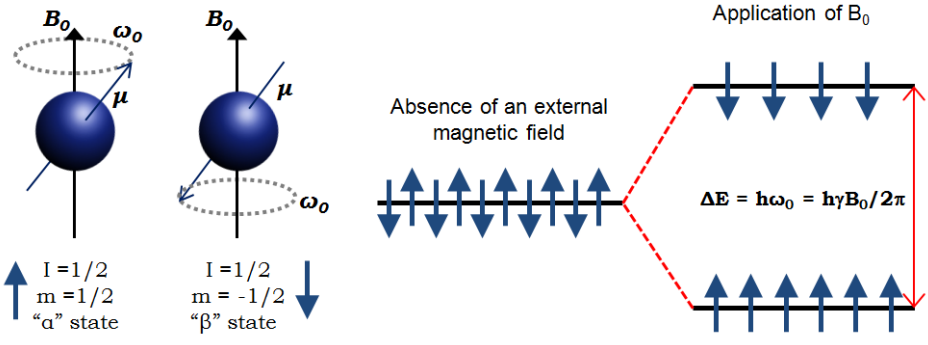


Figure 3.1. *Schematic representation of the two possible orientations of the magnetic moment for a nucleus with I = ½ that precesses with a Larmor frequency ω_0. On the right, the differentiation of both spin states in terms of energy after applying an external B_0.*

Where "h" is the Plank constant and "γ" is the so-called *gyromagnetic ratio*, a constant value characteristic for each particular nucleus (Figure 3.2). All these states are degenerate in energy and therefore indistinguishable a priori. However, under the influence of an external magnetic field (B_0), the degeneracy disappears and some levels are more stable in energy than others,

depending on their "m" values. Consequently, this hyperfine distribution of the spins becomes asymmetric since more nuclei will tend to preferentially occupy the lowest energy levels. NMR is aimed at promoting changes in the spin populations by means of electromagnetic pulses and observing the emission of radiofrequencies from those nuclei which are trying to reach the equilibrium again. The energy of each magnetic moment state is:

$$E = \mu_z \cdot B_0 = m\gamma B_0 \cdot \frac{h}{2\pi} \qquad (Eq.\,2)$$

For a dipolar nucleus with I = 1/2, the application of Eq. 2 allows to calculating the difference in energy between the two possible states (m_I = +1/2, m_I = -1/2) (Eq. 3). Such a difference corresponds to an electromagnetic frequency that turns out to be the precession frequency of the magnetic moment vector, also called the **Larmor frequency (ω_0):**

$$\Delta E = h \cdot \omega_0 = \gamma B_0 \cdot \frac{h}{2\pi} \quad \rightarrow \quad \boldsymbol{\omega_0 = \frac{\gamma B_0}{2\pi}} \qquad (Eq.\,3)$$

	Nuclear spin (I)	Natural abundance (%)	Gyromagnetic ratio ($/10^7$ rad T^{-1} s^{-1})	Resonant frequency (MHz)	Relative sensibility
^{1}H	1/2	99.98	26.75	500	1
^{2}H	1	0.02	4.11	76.75	0.00000145
^{13}C	1/2	1.11	6.73	125.72	0.000176
^{14}N	1	99.63	1.93	36.12	0.00101
^{15}N	1/2	0.37	-2.71	50.66	0.00000385
^{17}O	5/2	0.04	-3.63	67.78	0.0000108
^{19}F	1/2	100.00	25.18	470.39	0.83
^{31}P	1/2	100.00	10.84	202.40	0.0663

Figure 3.2. *Properties of some common active isotopes employed in NMR. Their resonant frequencies correspond to a field strength of 11.74 T, which confers the ^{1}H nucleus a Larmor frequency of 500 MHz. Adapted from Ref [34].*

Noteworthy, the ΔE term is remarkably low and the Boltzmann distribution predicts an incredibly small difference between the populations of both possible spin levels. Indeed, for a 400 MHz frequency, the gap created in the ^{1}H spin states leads to only a 0,0064% of the nuclei staying at the high energy state. This is the main reason for which NMR is commonly regarded as an insensitive technique, as most of nuclei in a sample do not generate observable transitions. Apart, the ^{1}H is the most sensitive atomic nucleus regarding its natural abundance (99,98%) and its particular gyromagnetic ratio, while for other nuclei, relative sensitivities are lower (Figure 3.2).

Excluding ^{1}H, the active isotopes of other relevant biological elements, mainly ^{13}C and ^{15}N, are very scarce in nature. Obviously, from Eq. 3 it can be deduced that the only way to achieve better resolutions (higher ω_0) is to increase the external magnetic field. Though, the magnetic field enhancement is not enough for ^{13}C and ^{15}N observation in many cases and isotopically enriched samples are often needed for the study of such nuclei in biomolecules (*see below*).

Without getting into detail, the work pipeline for performing an NMR measurement firstly requires the application of a set of radiofrequencies centered on a reference ω_0 value. Thus, the bulky magnetization of the sample (M_Z) is perturbed and subsequently treated with sequential pulses aimed at transferring the magnetization (coherence) between nuclei close in space and permitting certain "properties" of the perturbed nuclei to evolve in different ways. Finally, the system is allowed to relax, emitting a complex signal that decays along the time: the **Free Induction Decay (FID)** signal. Afterwards, this time-domain signal is transformed into another function in the domain of frequencies (Fourier Transform, FT), yielding what we know as an NMR spectrum.

Noticeably, an NMR experiment would not have sense whether all the nuclei are affected by the external field in the same way. Actually, the applicability of the NMR methodologies lies in the particular Larmor frequency that every spin displays depending on its chemical environment in a given molecule. This local influence is manifested through several of the aforementioned "properties" which each spin externalizes and whose nature elicits different noticeable effects. In this regard, the common NMR parameters describing these phenomena are the chemical shifts [35], the scalar couplings [36], the nuclear Overhauser effects (NOEs)[37] and the relaxation parameters:

- The electron cloud that surrounds every nucleus is responsible for the observed **chemical shifts**. Literally, the basic Larmor frequency arising from the applied B_0 is slightly tuned by additional small magnetic fields generated by the circulating electrons. These little fields oppose to the main B_0, giving rise to subtle variations (less than a kilohertz) on the ω_0 of the studied nuclei, originating the common signal dispersion observed around the reference frequency ($B_0 = h\omega_0$). Similarly, a reduction in the expected electronic density leads to the opposite effect. In the former case, the nuclei are said to be shielded (high field), while the contrary is true for the latter case.

The influence of the electron cloud strongly depends on how it is actually affected by the surrounding network of covalent and non-covalent bonds [34]. In general, the electronic perturbations transmitted through covalent bonds include inductive and mesomeric effects, both arising from asymmetric charge distributions originated at electron-withdrawing groups, electron-donating groups or ionic species. In biomolecules, like proteins or nucleic acids, the anisotropic effects are more relevant by far. These are chiefly caused by aromatic rings and carbonyl groups, unsaturated systems in which the circulation of the π electrons creates notable magnetic perturbations directly transmitted through space. Analogously, the presence of paramagnetic ions has similar effects with a remarkable strength: pseudocontact shifts (PCS) [38]. Particularly in the case of protons, chemical exchange from acidic or basic groups affects the chemical shift as well, as the signal is averaged with that from the solvent protons.

In practice, chemical shifts depend on the specific B_0 applied. Hence, the δ scale (in ppm) was created as an universal scale which always provides the same chemical shift value for a given nucleus regardless the external field:

$$\delta(\text{ppm}) = \frac{\omega_{\text{obs}} - \omega_{\text{ref}}}{\omega_{\text{ref}}} \cdot 10^6 \qquad (Eq. 4)$$

- As stated before, atomic nuclei genuinely behave as little magnets, exerting a noticeable influence on its neighboring atoms. The existence of two spin states necessarily implies that a close nucleus is sensing two different environments, what in turn leads to a signal splitting in two peaks, jointly describing the two possibilities (Figure 3.3). This mutual interference through covalent bonds is known as **scalar coupling,** while the difference in Hz between the resulting peaks is called coupling constant. These couplings merge as long as the two considered nuclei are not chemically equivalent (isochronous). Usually, noticeable coupling constants are established between atoms separated by two bonds (geminal, ^2J) or three bonds (vicinal, ^3J).

  Vicinal coupling constants [39] are of special interest in interaction studies due to its usefulness for many purposes as ring puckering in carbohydrates [40-41], rotamers and stereoisomers in proteins [42] or through H-bond couplings in nitrogen bases from nucleic acids [43]. Their

values are determined by several factors (substituents, bond lengths, bond valence angles), being the dihedral angle (ϕ) the factor underlying most of the mentioned applications. The dependence between ^3J and the dihedral angles was empirically established by Karplus at first [44], giving rise to the known Karplus curves that are widely used today for predicting torsional angles in biomolecules [45] (Figure 3.4). The parameters of the general equation (Eq. 5) depend on the particular nuclei pair studied:

$$^3\mathbf{J} = \mathrm{A} - \mathrm{B}\cos(\varphi) + \mathrm{C}\cos^2(\varphi) \qquad (Eq.5)$$

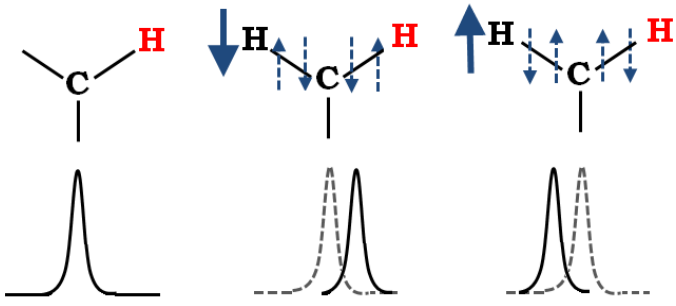


Figure 3.3. *The NMR signal given by an isolated proton (in red) compared with the signal that result from the presence of a neighboring nucleus with two possible spin states, whose influence is transmitted through the covalent bond network. The thick arrows represent the nuclear spins and the dashed arrows correspond to the electronic spins.*

$^3J(\mathrm{HNH}^{\alpha})$	$^3J(\mathrm{HNH}^{\alpha}) = 6.98\cos^2(\varphi\text{-}60°) - 1.38\cos(\varphi\text{-}60°) + 1.72$
$^3J(\mathrm{HNC}^{\beta})$	$^3J(\mathrm{HNC}^{\beta}) = 3.39\cos^2(\varphi\text{+}60°) - 0.94\cos(\varphi\text{+}60°) + 0.07$
$^3J(\mathrm{HNC'})$	$^3J(\mathrm{HNC'}) = 4.32\cos^2(\varphi\text{+}180°) + 0.84\cos(\varphi\text{+}180°)$
$^3J(\mathrm{C'}_{i-1}\mathrm{H}^{\alpha})$	$^3J(\mathrm{C'}_{i-1}\mathrm{H}^{\alpha}) = 3.75\cos^2(\varphi\text{+}120°) - 2.19\cos(\varphi\text{+}120°) + 1.28$
$^3J(\mathrm{C'}_{i-1}\mathrm{C}^{\beta})$	$^3J(\mathrm{C'}_{i-1}\mathrm{C}^{\beta}) = 1.59\cos^2(\varphi\text{-}120°) - 0.67\cos(\varphi\text{-}120°) + 0.27$
$^3J(\mathrm{C'}_{i-1}\mathrm{C'})$	$^3J(\mathrm{C'}_{i-1}\mathrm{C'}) = 1.33\cos^2(\varphi) - 0.88\cos(\varphi) + 0.62$
$^3J(\mathrm{H}^{\alpha}\mathrm{N}_{i+1})$	$^3J(\mathrm{H}^{\alpha}\mathrm{N}_{i+1}) = -0.88\cos^2(\psi\text{-}120°) - 0.61\cos(\psi\text{-}120°) - 0.27$
$^3J(\mathrm{H}^{\alpha}\mathrm{H}^{\beta 2})$	$^3J(\mathrm{H}^{\alpha}\mathrm{H}^{\beta 2}) = 9.5\cos^2(\chi^1\text{-}120°) - 1.60\cos(\chi^1\text{-}120°) + 1.8$
$^3J(\mathrm{H}^{\alpha}\mathrm{H}^{\beta 3})$	$^3J(\mathrm{H}^{\alpha}\mathrm{H}^{\beta 3}) = 9.5\cos^2(\chi^1) - 1.68\cos(\chi^1) + 1.8$
$^3J(\mathrm{H}^{\alpha}\mathrm{C}^{\gamma})$	$^3J(\mathrm{H}^{\alpha}\mathrm{C}^{\gamma}) = 7.1\cos^2(\chi^1\text{+}120°) - 1.0\cos(\chi^1\text{+}120°) + 0.7$
$^3J(\mathrm{C'H}^{\beta 2})$	$^3J(\mathrm{C'H}^{\beta 2}) = 7.2\cos^2(\chi^1) - 2.04\cos(\chi^1) + 0.6$
$^3J(\mathrm{C'H}^{\beta 3})$	$^3J(\mathrm{C'H}^{\beta 3}) = 7.2\cos^2(\chi^1\text{+}120°) - 2.04\cos(\chi^1\text{+}120°) + 0.6$
$^3J(\mathrm{NH}^{\beta 2})$	$^3J(\mathrm{NH}^{\beta 2}) = -4.4\cos^2(\chi^1\text{+}120°) + 1.2\cos(\chi^1\text{+}120°) + 0.1$
$^3J(\mathrm{NH}^{\beta 3})$	$^3J(\mathrm{NH}^{\beta 3}) = -4.4\cos^2(\chi^1\text{-}120°) + 1.2\cos(\chi^1\text{-}120°) + 0.1$

Figure 3.4. *Examples of empirical Karplus equations describing the dependence between the coupling constants from nuclei at the protein backbone and the corresponding torsional angles φ, ψ and χ. Adapted from [42].*

- Apart from scalar couplings, direct perturbations may appear between atoms that lie very close to each other in the space. These perturbations originate the so-called **dipolar couplings** [46]. However, such couplings depend on the relative orientation of both nuclei with respect to the external B_0, and hence, free tumbling of molecules in solution eventually averages all these effects, making them invisible in the NMR spectrum. On the contrary, when a given nucleus in a two-spin system is irradiated with its specific frequency, the system tries to reach relaxation by transferring the magnetization through dipolar coupling mechanisms (W_0/W_2, Figure 3.5A). As a result, the irradiated nucleus is no longer visible and the signal of the counterpart experiences an intensity change that is named the **Nuclear Overhauser Effect (NOE)** [47]. The quantification of this effect is fairly complex. Broadly speaking, the maximum magnitude of the NOE depends on the half of the difference between both gyromagnetic ratios, and its value rapidly decreases as the interatomic distance grows (Figure 3.5B). Moreover, the balance between the W_0/W_2 mechanisms determines the NOE sign. At the same time, the predominant mechanism is related to the correlation time (τ_C), which correspondingly depends on the molecular size. When W_0 is the major process, the NOE is negative and reaches its maximum magnitude (Figure 3.5B).

- Finally, the specific relaxation pathways followed by the excited atoms can be likewise exploited to obtain useful information, essentially regarding dynamic aspects of the studied system. There are two main relaxation mechanisms: longitudinal and transversal relaxation [48-49].

  The **longitudinal relaxation (T_1)** is often said to be enthalpically driven, since it essentially comes from the loss of energy promoted by the fluctuating magnetic fields present in the surrounding lattice, which contribute to drive the excited nuclei back to the relaxed state. In contrast, the **transversal relaxation (T_2)** is entropically driven due to the spin-spin interactions (chemical shifts, couplings) contributing to the loss of phase coherence in the resonating nuclei. This relaxation is then associated to the attenuation of the neat magnetization component existing in the transversal *xy* plane, according to the vectorial model (Figure 3.6). T_2 is generally faster than T_1, except for particular conditions that favor the latter mechanism, like the presence of radical species or quadrupolar nuclei ($I > 1/2$). Therefore, relaxation studies mainly focus on

monitoring the transversal relaxation component of the observed system (from an entire molecule or different parts of it).

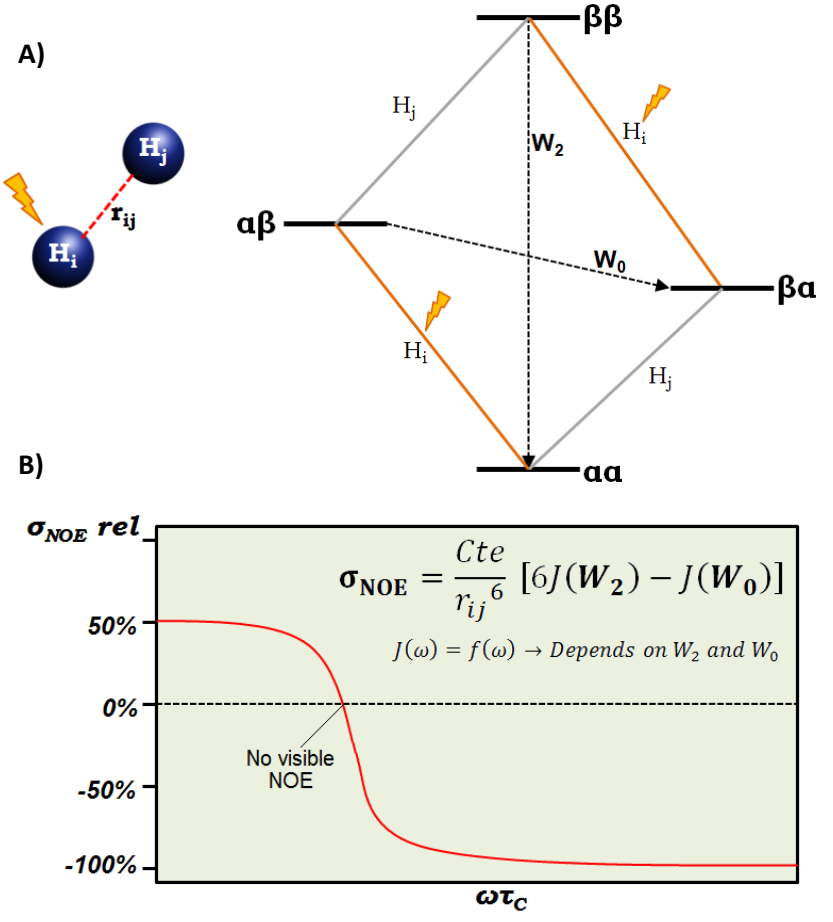


Figure 3.5. *A) Diagram of the existing energy levels for a two-spin system. If the nucleus labeled as H_i is irradiated, the relaxation pathways depicted as orange lines are disabled. To reach the ground state, the system can evolve through cross relaxation mechanisms enabled by dipolar couplings, represented as black dashed arrows (W_2 and W_0). B) The dependence of the NOE magnitude and sign on different parameters like the correlation time (τ_C) or the interatomic distance (r_{ij}).*

In the biological context, also including this work, the often large size of the studied entities (oligosaccharides, proteins) precludes the sole utilization of monodimensional (1D) spectra for gathering the desired data. Signal crowding is a typical problem in 1D-^{1}H NMR, whereas other nuclei like ^{13}C and ^{15}N offer a better spectral dispersion but entail severe sensibility problems when measured individually (low isotopic abundance, lower γ). Then, multidimensional NMR is meant to solve these problems by expanding the

NMR correlations to additional dimensions [50-51]. Commonly, the pulse sequences incorporate an additional evolution step with a variable delay which prepares the system for a subsequent stage in which other components of the spin-spin interactions can develop (mixing time). At last, the FID is routinely processed by FT and then, a second FT is applied to obtain a planar spectrum (2D) containing the information. For more dimensions, the FID processing is analogous (nFT) although the complexity increases. The resulting experiments are classified as homonuclear or heteronuclear correlation depending on whether the correlated nuclei are respectively identical (2D-COSY, 2D-TOCSY, 2D-NOESY…) or different (2D-HSQC, 2D-HMQC, 2D-HMBC, 3D-HSQC-NOESY, 3D-HNCA…) [19].

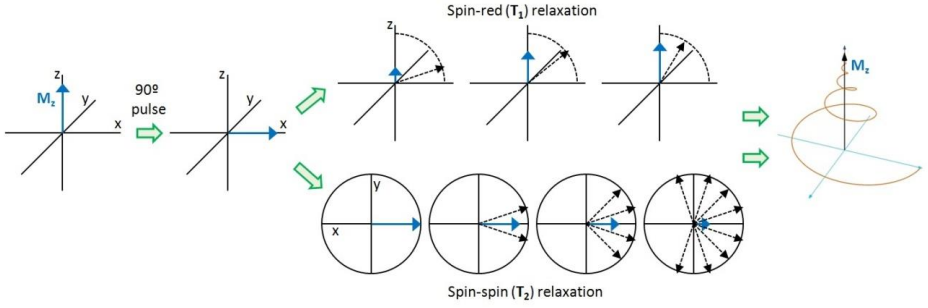


Figure 3.6. *The two relaxation mechanisms that contribute to take the neat magnetization of the nuclei (M_Z) from the xy plane back to the z axis.*

Regarding protein-ligand interactions, there are two habitual strategies considered when making use of NMR approaches. The interaction can be independently studied from the point of view of each entity, analyzing how the recognition process affects their respective resonances [52-54]:

- The *receptor-based methods* (in our case, the DC-SIGN CRD or ECD) are based on monitoring the perturbations on the protein NMR resonances occurring during the binding event. This method allows to delineating the binding site and differentiates between specific and non-specific interactions. In addition, there is not an affinity limit for the measurements and it is possible to observe interaction kinetics and thermodynamics. The most employed methodologies include **HSQC-based titrations**, **relaxation (^{15}N, ^{13}C) and diffusion editing (2D-DOSY).** By contrast, the typical drawbacks associated to these methods arise from the size limit (< 40 kDa) [55]. Large receptors possess very efficient relaxation mechanisms (T_2) and display profound signal broadening. Moreover, even having a good spectrum quality, the intense signal

overlapping necessarily demands the use of multidimensional NMR (2D, 3D, 4D), what requires the use of large amounts of the receptor and isotopically labeled samples (^{13}C, ^{15}N) whose preparation is often expensive and time-consuming.

- *Ligand-based methods* are aimed at extracting information from the ligands in the presence and absence of the receptor. A crucial advantage of these methods reside in overcoming the cited limitations regarding the receptor: in contrast to receptor-based methods, the molecular size is not limiting, even some experiments actually yield better ligand signals as the receptor is larger (e.g. STD). Additionally, the required amounts of receptor are considerably lower and isotopic labeling is not needed. Typical examples of such experiments are **Saturation Transfer Difference (STD), Transferred-NOESY (TrNOESY), Water-LOGSY and relaxation (^{19}F, ^{13}C)**. The extracted data allow to detecting ligand binding (ligand screening) as well as estimating affinities, in some cases. However, these methods usually rely on the sufficient difference in size between receptor and ligand, whereby the study of large ligands may entail measuring limitations. Similarly, the ligand is often observed in its free state after leaving the receptor, and then the study of high-affinity ligands ($K_d < 10^{-7}$ M) is hampered by their intrinsic low k_{off} rates.

In the next sections, the main practical aspects about the experiments used in this work are detailed. ^{15}N-HSQC-based titrations have been employed for the lectin analysis (CRD), while the oligosaccharides of interest (Blood Group antigens, monosaccharides) have been fundamentally analyzed by STD, TrNOESY and ^{19}F-T2 relaxation (with fluorinated sugars). Also, the analysis of the CRD required the previous assignment of the protein signals by three-dimensional NMR, as specified below.

3.2. NMR assignment of the DC-SIGN CRD

Proteins are the extreme case of signal crowding in routine 1D-^{1}H NMR spectra. A single protein having 100 amino acid residues or more readily contains several hundreds of protons dispersed along 12 ppm as much. Thus, assignment of proteins is unworkable by only considering a single dimension. In the middle eighties, the Nobel Prize winner Kurt Wüthrich managed to solve by the first time the three-dimensional structure of a globular protein through NMR methodologies [56]. The Wüthrich methodology (known as “sequential assignment”) is based on recording two kinds of experiments: 2D-TOCSY and

2D-NOESY. The 2D-TOCSY (TOtal Correlation SpectroscopY) resembles the COSY experiment, but it permits to transfer the magnetization to all the protons connected by scalar couplings [57]. In the context of a protein, these connections remain isolated within each individual amino acid. Then, a specific correlation pattern is meant to fit with a specific sidechain, hence allowing the classification of the observed patterns by residue type [58] (Figure 3.7A). In principle, each amino acid will give at least one TOCSY pattern, except for those sidechains containing other isolated spin systems, for instance in aromatic residues (Phe, Tyr, Trp, His). As complement, the 2D-NOESY provides the spatial connections between close nuclei, enabling the sequential correlation between neighboring residues. Regardless the three-dimensional structure of the protein or peptide studied, some sequential NOEs always appear, as those between $H\alpha_i$ and HN_{i-1} [59-60] (Figure 3.7B).

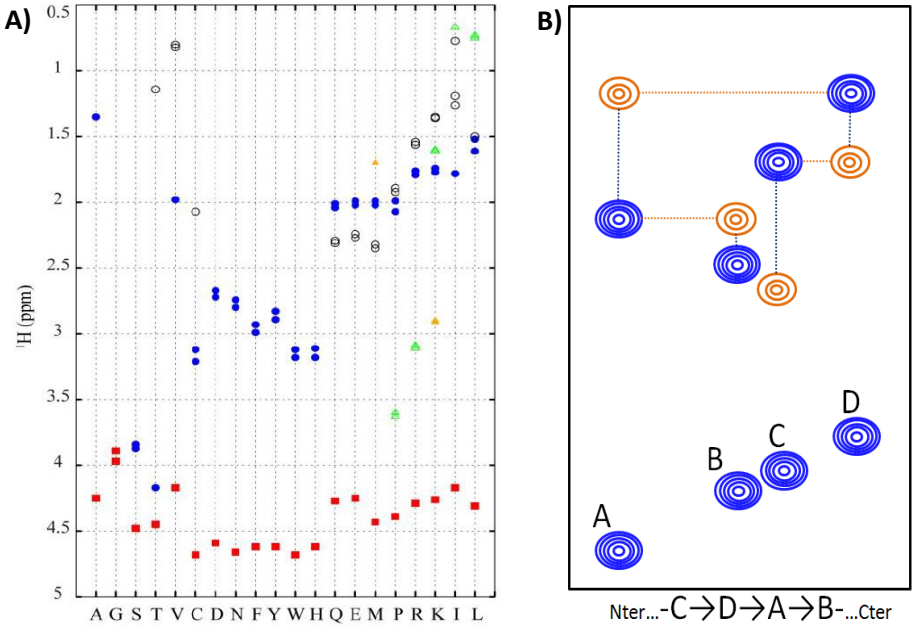


Figure 3.7. *A) Average values of proton chemical shifts in proteins according to the BMRB database [58]. Red squares: Hα. Blue circles: Hβ. Empty black circles: Hγ. Green triangles: Hδ. Orange triangles: Hε. B) Sequential correlations established through NOE effects observed in the amide region of the 2D-NOESY spectrum (6-9 ppm). In blue, the vertically derived TOCSY correlations. In orange, the NOE crosspeaks between HN(i) and Hα(i-1).*

In protein-ligand studies, the reference receptor-based method is the HSQC-based titration [61]. The HSQC experiment (Heteronuclear Single-Quantum Correlation) is a 2D spectrum showing the correlation between proton and another nucleus, mainly ^{15}N. The ^{15}N-^{1}H-HSQC shares the same principle of

"residue isolation" as the 2D-TOCSY when applied to proteins: the direct H-N correlations essentially come from the amide bonds, and thereby each amino acid provides a single peak corresponding to its HN group. As exceptions, Trp, Asn and Gln do exhibit two HN peaks, with the additional one belonging to their particular sidechains. Similarly, His and Arg have exchangeable HN protons that occasionally appear under certain conditions (acidic). In contrast, Pro is devoid of amide protons in protein structures and therefore becomes invisible in the ^{15}N-HSQC.

Although effective for small peptides and proteins, the classical method of Wüthrich eventually fails as the protein size increases over 80-100 residues, because signal overlapping becomes extremely severe even in two dimensions. Three-dimensional spectra constitute the most practical solution in such cases [62]. Following the same methodology, the TOCSY and NOESY patters can be extended to a third dimension by referencing them to their respective HN groups in the ^{15}N-HSQC, giving rise to experiments like 3D-^{15}N-HSQC-TOCSY and 3D-^{15}N-HSQC-NOESY. However, the sensitivity of these experiments is deeply decreased by dipole-dipole relaxation affecting the homonuclear transference of coherence in large structures. Better, the recent development of pulse sequences based on *triple resonance* (^{1}H, ^{13}C, ^{15}N) has given rise to a plethora of different spectra correlating vicinal carbon nuclei with amidic HN groups [63-64]. The configuration of the pulse sequence determines which carbons (Cα, Cβ, CO, Cγ…) receive the magnetization. The key factor is that this transference of magnetization can be directed either to the proper sidechain carbons or to carbons from vicinal residues, thereby providing very few and clean signals which enable the unambiguous correlation of sequential residues [65]. For the particular work described here, four distinct experiments have been employed to establish such sequential correlations and identify as many residues of the DC-SIGN CRD as possible. These experiments are dubbed HNCA [66], HN(CO)CA, HN(CO)CACB [67] and HNCACB [68-69] (Figure 3.8).

The HN(CO)CA experiment gives an only correlation between an HN_i and the alpha carbon from the previous residue, $C\alpha_{i-1}$. Similarly, the HNCA experiment shows two correlations in the carbon dimension, both with carbons at position α: $C\alpha_{i-1}$ (the previous residue) and $C\alpha_i$ (the same residue). Given that the $C\alpha_{i-1}$ can be readily assigned in the HN(CO)CA spectrum, the remaining assignment ($C\alpha_i$) is immediate. Theoretically, these assignments could be enough to establish sequential assignments. The main problem arises from the low chemical shift dispersion displayed by Cα nuclei, habitually varying within a 15

ppm range (50-65 ppm) and being Gly the only exception (43-48 ppm) [70]. This reduced range often leads to many ambiguities, whereby two additional experiments, HN(CO)CACB and HNCACB, are likewise performed to add a second carbon nucleus (Cβ) as an exclusion rule for those ambiguous assignments. Besides a slightly higher spectral window (24-48 ppm), Cβ provides a great way to immediately identify four amino acid types: Gly (No Cβ), Ala (15-21 ppm), and Ser & Thr (very deshielded, 60-75 ppm). Specifically, the HN(CO)CACB displays two carbon correlations: $C\alpha_{i-1}$ and $C\beta_{i-1}$. The $C\beta_{i-1}$ resonance is unequivocally assigned as the $C\alpha_{i-1}$ is already identified in the HN(CO)CA. Eventually, the HNCACB is the most complete experiment, giving four carbon correlations for each single HN crosspeak: $C\alpha_{i-1}$, $C\beta_{i-1}$, $C\alpha_i$ and $C\beta_i$. Only the $C\beta_i$ resonance needs to be assigned, as the other three nuclei are provided by the experiments mentioned above.

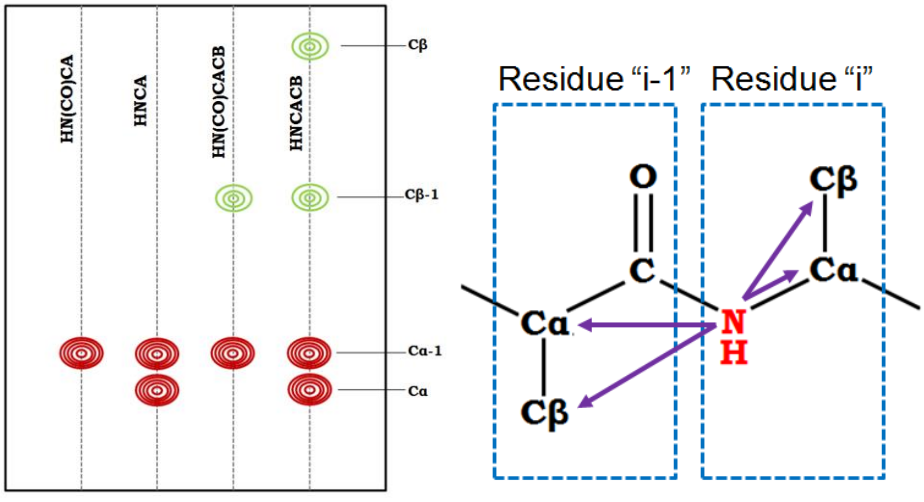


Figure 3.8. *Interresidue correlations deduced from the 3D NMR experiments HN(CO)CA, HNCA, HN(CO)CACB and HNCACB. The pulse sequences utilized are usually edited to show Cα and Cβ with opposite sign and ease their identification.*

Following this protocol, a complete assignment can be achieved by stepwise linking successive residues via the Cα and Cβ resonances. This sequential analysis may be uniquely stopped by prolines or highly flexible regions where undesired relaxation processes erase most of the signals. However, the recording of alternative available experiments (HNCO, HNCACO) in parallel might help to overcome these problems.

As complementary information, chemicals shifts have been shown to be good descriptors of secondary structure elements in proteins [71]. Variations in the proton chemical shifts are predominantly originated by the aforementioned

anisotropic effects and the hydrogen bonds. Instead, chemical shifts of Cα and Cβ strongly depend on the backbone torsion angles affecting the amino acid itself (ϕ, ψ) rather than through-space effects. On this basis, it has been noticed that Cα and Cβ resonances are good parameters to predict the presence of secondary structures (α-helix, β-strand) [72]. Their chemical shifts systematically change with respect to the expected values in "pure conditions", like in solvent-exposed random coil peptides [73-74], without considering the existence of a folded scaffold. Through computational tools, like TALOS+ [75] or PREDITOR [76], the secondary structural elements can be predicted with reasonable accuracy jointly computing δCα, δCβ and also δHα and δHN. The same chemical shifts serve as indicators of the local flexibility as well: the higher the flexibility is, the closer the chemical shifts are to those exhibited in random coil peptides. Chemical shift processing is also accomplished by computational methods, like the Random Coil Index (RCI) server [77].

3.3. ^{15}N-HSQC-based titrations

Amongst the receptor-based methods, the analysis of perturbations in the 2D-HSQC spectrum is the most representative technique, broadly applied to map the binding site, analyze the nature of the interaction (specific/non-specific [78], cooperativity [79], binding stoichiometry [80]…), control protein dynamics, estimate dissociation constants or even for ligand screening in drug design (Structure-Activity Relationship -> SAR by NMR [81-82]). In spite of the limitations regarding the molecular size, the continuous development of new hardware and pulse sequences has led to expand the biological scope of NMR to more complex receptors. Among these advances, the most relevant in this regard are optimized HSQC sequences like TROSY (Transverse Relaxation Optimized SpectroscopY)[83], which yield better resolution spectra; improved cryo-probe technologies [84]; perdeuteration of large proteins to avoid relaxation issues [85-86] and non-uniform sampling (NUS) [87], which substantially reduces the time costs without affecting the resolution.

In practice, two kinds of spectra are viable to perform these analyses: ^{13}C-HSQC and ^{15}N-HSQC. The choice may depend on the type of system and its molecular size, but always looks for those nuclei efficiently sensing the ligand approaching [88-89]. From an empirical perspective, ^{15}N, ^{13}Cα and ^{1}Hα resonances are the most affected by spatial factors like hydrogen bonding or near functional groups (for instance, from the ligand) causing local perturbations on the effective B_0 [90]. In contrast, amide ^{1}HN and ^{13}Cβ remain

less altered or its variation depends almost exclusively on conformational changes (Cβ) or local hydrogen bonding (^{1}HN). At last, the ^{1}H-^{15}N-HSQC is the most used for practical considerations (major amount of data in the shortest time). The ^{13}C-HSQC of proteins displays rather higher signal overlapping problems that are mostly reduced by means of selective ^{13}C-labeling [91-92], commonly applied to methyl groups in aliphatic residues (Leu, Val, Ile, Met). This methodology is better employed to carry out recognition studies involving very large receptors [93-94].

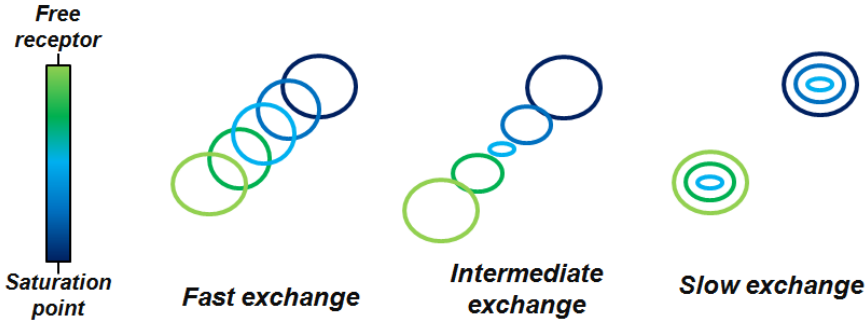


Figure 3.9. *Evolution of crosspeak shifts and shapes along a titration process depending on the type of exchange regime in the NMR timescale.*

The primordial application of the ^{15}N-HSQC is for depicting the binding site by detecting those amino acids that are the most influenced by the presence of the ligand. Nevertheless, an important consideration must be taken into account, which arises from binding kinetics: two different scenarios may show up two types of chemical shift drift [61]. If the chemical exchange between the free and the bound states takes place faster than the acquisition time, then we are facing a *fast exchange regime* in which the NMR signal from both states (free/bound) is averaged to a single crosspeak that progressively moves across the spectral window [95]. On the contrary, if the acquisition time is faster than the time the ligand needs to leave the binding site (tight binding), then the system is said to be in a *slow exchange regime*, and we will see how the free signal gradually vanishes as the ligand concentration grows, while a new signal (bound state) independently appears at another frequency [96] (Figure 3.9). In the limit between both kinds of events (*intermediate exchange*), peaks shift and broaden simultaneously [97]. In terms of kinetics, the comparison between the exchange rate (k_{ex}) and the chemical shift difference (in Hz) between free and bound states permits to distinguish the specific type of exchange regime. k_{ex} is directly correlated with the kinetic constants through the expression on Eq. 7 [52,98]. To consider, the second

term of this equivalence ($\pi\Delta\omega/\sqrt{2}$) can be uniquely included when the system is near to a regime of intermediate exchange, in which the middle points of the titration show the maximal broadening, even disappearing [99]. This phenomenon is called *"peak coalescence".* Usually, the exchange regime is directly correlated with k_{off} rather than with k_{ex}. In general, large k_{off} values (> 10^5 s^{-1}) correspond to fast exchange events and vice versa (slow exchange for k_{off} < 10 s^{-1}).

$$P + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PL$$

$$\text{At the equilibrium: } k_{on}[P][L] = k_{off}[PL] \rightarrow \mathbf{K_d} = \frac{\mathbf{k_{off}}}{\mathbf{k_{on}}} = \frac{\mathbf{[P][L]}}{\mathbf{[PL]}} \quad (Eq.6)$$

Where K_d is the dissociation constant at the working temperature and k_{on} and k_{off} are the corresponding kinetic constants for the formation and dissociation of the complex, respectively.

$$\text{At the coalescence temperature: } \mathbf{k_{ex}} = \boldsymbol{k_{on}[L] + k_{off}} = \frac{\boldsymbol{\pi \cdot \Delta\omega}}{\sqrt{\mathbf{2}}} \quad (Eq.7)$$

Where $\Delta\omega$ is the difference in Hz between the starting point (free receptor) and the end position (bound receptor).

For evaluating the chemical shift perturbations (CSP, $\Delta\delta$), both the ^{1}H CSPs and the ^{15}N CSPs are treated within the same mathematical expression to calculate an averaged chemical shift perturbation value [100-101], which is then taken as reference for deciding whether the variation is significant or not (Eq. 8). Obviously, the weighing of each nucleus (α_i) is different as its behavior upon ligand binding is so as well. Considering the proton as reference (α_H = 1), the α_N factor can take different values depending on how the ^{15}N perturbation is compared with that of the ^{1}H. Experimental values in literature commonly go from α_N = 0.1 [102] to α_N = 0.2 [103]. α_N = 0.1 comes from strictly using Hz instead of ppm ($\gamma_{15N} = 0.1\gamma_{1H}$). However, empirical observations have led many researchers to agree that the most suitable value is around α_N = 0.14 [104], considering that for the amide groups in proteins, the spectral window of the ^{15}N is approximately seven times wider than that of ^{1}H (1/7 ≈ 0.14). In this thesis, the standard criterion has been also adopted and a value of α_N = 0.14 has been applied.

$$\Delta\delta_{ave} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\alpha_i \Delta\delta_i)^2} \xrightarrow{For\ H/N} \Delta\boldsymbol{\delta_{HN}} = \sqrt{\frac{\mathbf{1}}{\mathbf{2}}[\Delta\boldsymbol{\delta_H}^2 + (\boldsymbol{\alpha_N}\Delta\boldsymbol{\delta_N})^2]} \quad (Eq.8)$$

Regarding the threshold value, there is neither a specific rule, although the typical cutoff values for deciding the significance of the CSPs are the standard deviation, σ, or twice the standard deviation, 2σ [105]. In the fast exchange regime, the resulting Δδ is contributed by the corresponding molar fractions of the receptor in their bound and free forms, respectively. Thus, the affinity constants can be deduced from the binding isotherm built by plotting the CSP as a function of the total ligand concentration [106]:

$$\delta_{obs} = \delta_F x_F + \delta_B x_B \ \rightarrow \ \delta_{obs} = \delta_F(1 - x_B) + \delta_B x_B \ \rightarrow$$

$$\delta_{obs} - \delta_F = x_B(\delta_B - \delta_F) \ \rightarrow \ \Delta\boldsymbol{\delta_{obs}} = \mathbf{x_B}\Delta\boldsymbol{\delta_{max}} \qquad (Eq.9)$$

Where x_B is the molar fraction of the bound receptor:

$$x_B = \frac{[PL]}{[P_0]} \ and \ [PL] = \frac{([P_0] - [PL])([L_0] - [PL])}{K_d} \xrightarrow{Therefore}$$

$$\Delta\boldsymbol{\delta_{obs}} = \frac{\Delta\boldsymbol{\delta_{max}}}{\mathbf{2}[\mathbf{P_0}]} \cdot \left(\mathbf{A} + [\mathbf{L_0}] - \sqrt{(\mathbf{A} + [\mathbf{L_0}])^2 - \mathbf{4}[\mathbf{P_0}][\mathbf{L_0}]}\right) \qquad (Eq.10)$$

And $A = [P_0]+K_d$. $[P_0]$ and $[L_0]$ are the total concentrations of receptor and ligand, respectively, being $[L_0]$ the variable for the exponential fit. The parameter $\Delta\delta_{max}$ is the chemical shift difference between the bound and the free state, and δ_B is taken from the observed chemical shift when receptor saturation has been reached ($[P_0]\approx[PL]$). Implicitly, the parameter A contains the value of the dissociation constant and the resulting plot is an exponential curve for a simple 1:1 binding event. For more complex binding events (induced fitting, multiple binding…), K_d cannot be accurately calculated since the CSPs suffer from other phenomena like non-linear drift [107-108] or uncommon broadening processes [109] (Figure 3.10).

For interactions occurring within the slow exchange regime, the CSPs cannot be collected and instead the intensity variations on the peaks of the free/bound states provide a similar relationship to estimate the dissociation constants [110]. In the intermediate regime, Eq. 10 may be employed as well, but being aware that the K_d estimation is expected to be worse as the dynamic behavior is less predictable, sometimes leading to more "sigmoidal" trends rather than ideal Langmuir-type curves [111] (Figure 3.10).

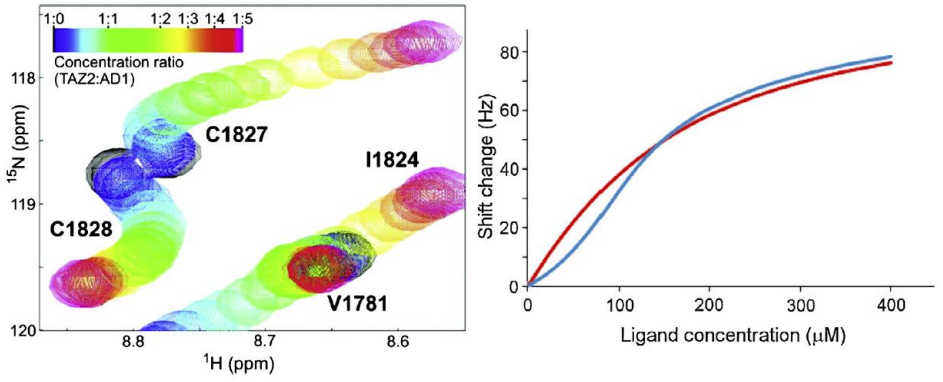


Figure 3.10. *On the left, curved trajectories observed for the stepwise recognition of several ligand molecules at different binding sites distributed on receptor´s surface (multiple binding). On the right, the distorted exponential growth (blue sigmoidal curve) observed when a binding process approaches to a regime of intermediate exchange. Adapted from [61,107].*

3.4. Saturation transfer difference (STD)

The STD experiment was originally conceived as a technique for rapid ligand screening using mixtures of different compounds [112]. It takes advantage of the different relaxation properties existing between a small ligand and its large receptor. In large proteins, the mechanisms of cross relaxation that originate the NOE effects are particularly efficient due to the extensive network of nuclei that are coupled through space. Thus, a selective pulse applied on a given frequency range will affect a specific group of protein protons that will become saturated and hence attenuated. By cross relaxation (spin diffusion), this saturation is transferred to the nearby protons and propagated throughout the entire protein structure [113].
Under these conditions, the proton signals of a compound that enters into the receptor binding site are susceptible to be also saturated, in this case through dipolar couplings transmitted by intermolecular mechanisms. On this basis, the STD setup is performed as follows (Figure 3.11A):

Firstly, a ^{1}H-1D spectrum is recorded while applying a train of selective pulses focused on a ppm region in which the protein exhibits NMR signals but the ligand does not. During this period, called *saturation time* (t_{sat}), the ligand molecules attached to the binding site get saturated. After they have been released from the recognition site, the magnetization status is still preserved and the free molecule displays weakened peaks as a result. This experiment is named **on-resonance** spectrum.

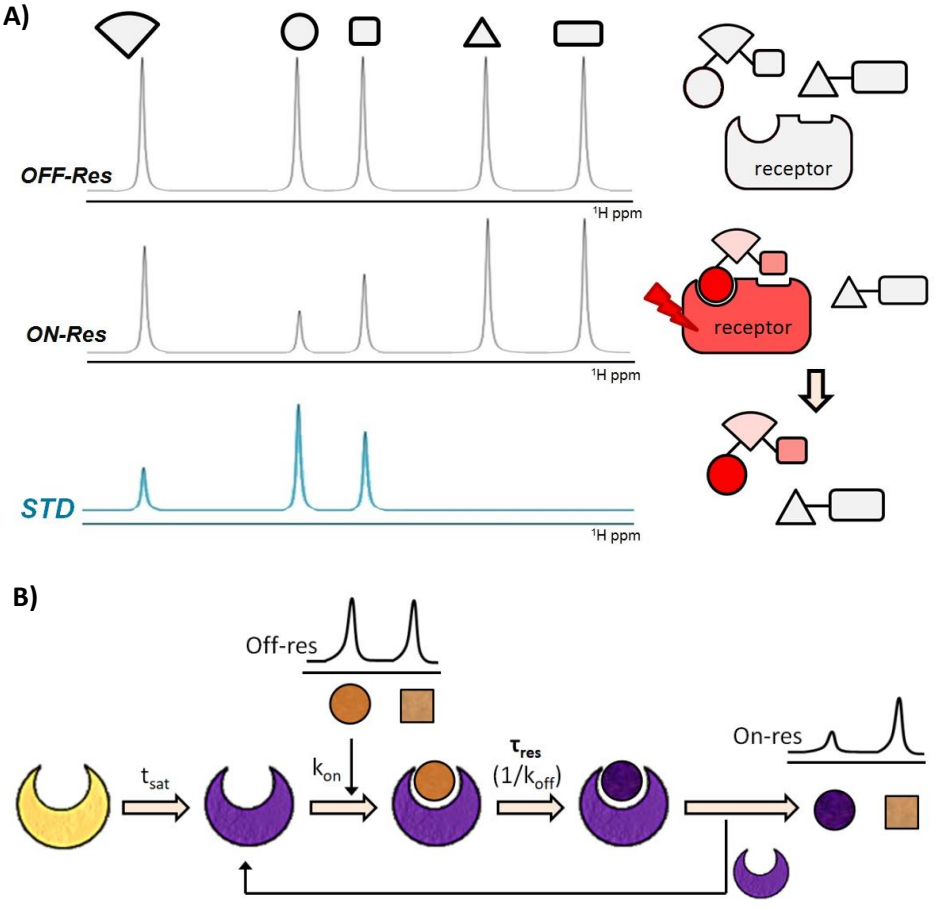


Figure 3.11. *A) The three individual spectra that constitute a complete STD experiment. On the top, the off-resonance spectrum, in which two ligands (structures on the right) show their corresponding proton signals. On the middle, the on-resonance spectrum displaying attenuated peaks for that ligand recognized by the saturated receptor. The attenuation for each moiety is different according with its distance to the protein (highlighted in red). On the bottom, the STD spectrum obtained as the difference between the two previous ones. The circle moiety displays the strongest STD as it is the closest fragment to the protein and thereby the most affected. B) The process of saturation transference that enables the acquisition of an STD: after the protein gets saturated (t_{sat}), the ligand needs to stay at the binding site for a specific time (τ_{res}) to be saturated. Eventually, after complex dissociation, the ligand still remains saturated if the dissociation rate is fast enough in the relaxation time scale, and exhibits sharp signals partially reduced due to that condition.*

In parallel, another ^{1}H-1D spectrum (blank) is obtained with the same pulse sequence, but the selective pulses are applied outside of the spectral window

(for instance, at 100 ppm). Thus, no signals are saturated at all, which is equivalent to record a routine 1D spectrum. This second spectrum is known as the **off-resonance** experiment and serves as reference. In the practice, the decrease on the signal intensities is very subtle and hence difficult to detect by directly comparing both spectra, on- and off-resonance. Therefore, a third spectrum is built by subtracting the on-resonance experiment to the reference ^{1}H-1D: the **STD spectrum.** As a result, the unperturbed signals totally disappear as their intensities are equivalent in both on- and off-spectra, while the affected peaks in the on-resonance spectrum are amplified even more as their intensities get weaker by transferred saturation. Remarkably, the adaptability of the pulse sequence enables the combination of this experiment with additional evolution steps, for instance allowing homonuclear couplings through a TOCSY scheme [114-116] or heteronuclear couplings through an HSQC sequence [117]. In turn, such versatility has enormously expanded its applications not only for ligand screening [118-119], but also in the study of ligand-receptor interactions from an atomic perspective [120].

With these features, the STD experiment constitutes a highly sensitive and robust tool to detect ligand binding [121]. Indeed, it efficiently avoids the detection of false positives since the compounds that do not interact with the saturated receptor are completely erased from the final STD. Noteworthy, the experiment relies on getting the highest possible percentage of saturated ligand molecules, while the bulky saturation is progressively lost in the free state by normal relaxation (T_1/T_2). Thus, the optimal scenario to record an STD demands the binding process to take place in a regime that is slow enough for the ligand to stay in the bound state the sufficient time (residence time, τ_{res}) to become saturated, but fast in the relaxation time scale to keep the saturation information in the free state during the acquisition [54] (Figure 3.11B). Under these requirements, the STD methodology has been demonstrated to be suitable for studying recognition processes with K_d values in the **10^{-3}-10^{-8} M range** [122]. In the practice, good STD sensitivities depend on the amount of free saturated ligand molecules. Considering fast exchange regimes, this condition enables the utilization of large protein/ligand ratios, usually ca. 1:50, to get intense enough STD spectra, thus considerably reducing the amount of receptor needed down to 1 μM. In addition, STD methodologies are well suited for studying ligand recognition processes driven by huge receptors: the larger the receptor is, the more efficient the transference of saturation is by NOE mechanisms and the stronger the STDs will be. Hence, STD protocols have been successfully carried out to investigate the recognition of small

compounds by very large systems, including liposomes [123] or viral particles [124].

Additionally, the applicability of the STD has been quickly extended over the years to two additional tasks: **ligand epitope mapping** [125-126] and **K_d estimation** [53,127]. Apart from assessing ligand attachment, it is possible to map the binding epitope in a given compound due to the different closeness of the ligand groups to the receptor surface [128-129]. Thus, cross relaxation will better affect those protons lying nearer to the receptor ones. However, special care must be taken when STDs are applied to this purpose. The STD intensities are regulated by many factors besides spatial distribution, being the most relevant the k_{off}, rebinding effects [130] and the individual T_1 relaxation of each proton [131].

Experimentally, the longitudinal relaxation has a different impact on each atom of a given molecule. Consequently, the initial saturation gained by a particular proton evolves such that its relative intensity might change after a long saturation time. The T_1 relaxation is therefore the main reason behind a biased interpretation of the STDs, which may lead to an incorrect epitope mapping. One good solution can be afforded by using STD measurements at short saturation times (around 0.25-0.50 s), although the amount of saturated molecules is reasonably lower and the sensitivity would be expectedly worse. Strictly, the only way to efficiently minimize the bias effect of the T_1 evolution is to use the so-called approximation of the *initial slopes*, based on plotting the STD evolution of each proton as a function of the saturation time (Eq. 11) and obtaining the slope for the exponential function when t_{sat} approaches zero [132]:

$$\mathbf{I_{STD}} = \mathrm{I_{MAX}} \cdot \left(1 - \mathrm{e}^{\mathrm{k_{sat}}\mathbf{t_{sat}}}\right) \qquad (Eq.\,11)$$

$$\frac{d(I_{STD})}{d(t_{sat})} = \mathrm{I_{MAX}} \cdot \mathrm{k_{sat}} \cdot \mathrm{e}^{\mathrm{k_{sat}}\mathbf{t_{sat}}} \xrightarrow{t_{sat}\,=\,0} \frac{\boldsymbol{d(I_{STD})}}{\boldsymbol{d(t_{sat})}} = \mathbf{I_{MAX}} \cdot \mathbf{k_{sat}} \qquad (Eq.\,12)$$

Typical saturation times range between 0.25 s and 4 s as much. Larger t_{sat} do not further improve the STD intensities as the saturation of almost all the ligand molecules has been nearly reached. The slope at $t_{sat} \approx 0$ is given by the product of the maximum achievable STD intensity (I_{MAX}) and the constant saturation rate (k_{sat}), as stated in Eq. 12. At the same time, the latter parameter allows to evaluating the relaxation properties of each proton and

predicts their STD build-up speeds. By direct comparison of k_{sat} values, it is possible to determine how a long saturation time will affect to the relative STD profile and therefore prevent erroneous interpretations.

Regarding the calculation of dissociation constants, a quantitative STD-based approach is commonly applied that requires a different treatment of the experimental dataset. Mayer and Meyer [133] firstly proposed the *amplification factor (STD_{AF})* as a new parameter to better assesses the absolute magnitude of the STDs. This value consists of multiplying the STD effect by the molar excess of the ligand, and thereby the measured STD is exclusively determined by the fraction of bound protein [134-135]:

$$\mathrm{STD_{AF}} = \frac{\mathrm{I_0 - I_{sat}}}{\mathrm{I_0}} \cdot \frac{[\mathrm{L_0}]}{[\mathrm{P_0}]} \qquad (Eq.\,13)$$

Being L_0 and P_0 the total amounts of ligand and protein, respectively, and I_0 and I_{sat} the corresponding peak integrals measured in the off- and on-resonance experiments. Using this parameter, the build-up curves are represented as a function of the total ligand concentration rather than the saturation time. The fitting equation is analogous to that used for describing enzymatic processes through the Michaelis-Menten model [136]. To note, such a model is an approximation that becomes valid only if $[L]_{free}$ approaches the total amount of ligand, $[L_0]$, what normally happens in STD experiments.

$$\mathrm{STD_{AF}} = \mathrm{STD_{AF,MAX}} \cdot \frac{[\mathrm{L}]}{\mathrm{K_d} + [\mathrm{L}]} \qquad (Eq.\,14)$$

Nevertheless, the K_d values have been checked to easily deviate from the real ones when using different saturation times, protons with distinct intrinsic STD effects or even different protein concentrations, in spite of using the same protein/ligand ratios [137-138]. Fortunately, the solution is fairly similar to that described above to evaluate the ligand epitope, but using the amplification factors at t_{sat} close to zero (STD_{AF0}) instead of the common STD effects. Briefly, the Langmuir-like isotherms are built by plotting the STD_{AF} versus the saturation time, using different ligand concentrations. For each $[L_0]$ value, its corresponding STD_{AF0} is inferred from the initial slope obtained through the same exponential equation (Eq. 11). Subsequently, the amplification factors at $t_{sat} \approx 0$ are correlated with the increasing ligand concentrations by means of Eq. 14 again (Figure 3.12). Definitely, ligand competition experiments constitute a fairly faster alternative, which involves the use of another known ligand or inhibitor whose dissociation constant (or

inhibition constant) has been already determined through other techniques [139]. In such a case, the unknown dissociation constant can be better approximated through a more complex mathematical expression, although some authors have also proposed the Cheng-Prusoff equation as a good estimation method [140]:

$$K_d = \frac{[L_0]K_I}{IC_{50} - K_I} \qquad (Eq. 15)$$

Where IC_{50} is the amount of inhibitor needed to reduce the intensity of the STD peaks from the unknown ligand "L" to the half. K_I is the known dissociation constant for the probe compound. The inhibitor is titrated in the presence of the protein-L system at a fixed concentration L_0.

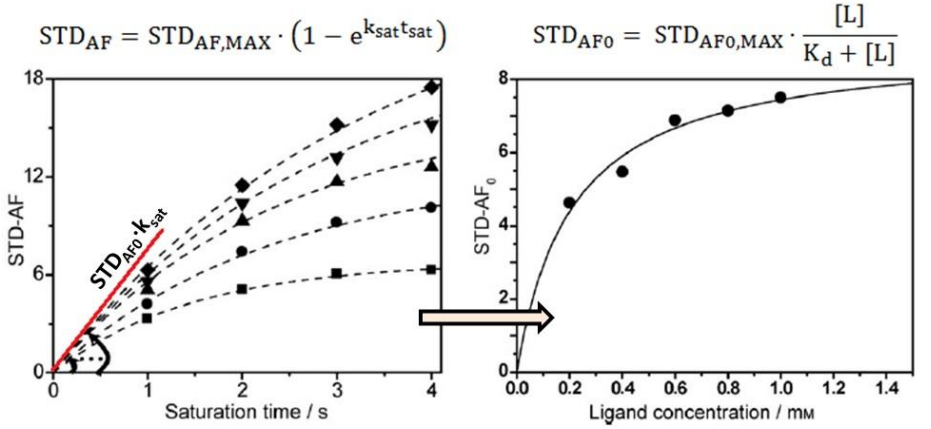


Figure 3.12. *Quantitative assessment of dissociation constants by STD NMR. The amplification factors are formerly recorded at different saturation times and ligand concentrations (dashed curves). For each [L], its corresponding STD_{AF} at t_{sat} = 0 is calculated through Eqs. 11 and 12. Then, the STD_{AF0} values are plotted as a function of the ligand concentration (right graph) using Eq. 14, from which the K_d can be finally deduced. Adapted from [137].*

Finally, it is worth mentioning the CORCEMA-ST software [141], which is capable of quickly assessing the veracity of theoretical structural models by using the experimental STDs as computing parameters. The theoretical STD values calculated with this tool are compared to the experimental ones through a statistical parameter called the NOE R-factor, which evaluates the degree of fitting between both datasets [142]:

$$R_{NOE}\ factor = \sqrt{\frac{\sum W_k (STD_{exp,k} - STD_{cal,k})^2}{\sum W_k (STD_{exp,k})^2}} \qquad (Eq.16)$$

Where STD_{exp} and STD_{cal} are respectively the experimental and theoretical STD effects. W_k is a weighting factor specific for each "k" proton, proportional to 1/ STD_{exp}. Ideally, if the theoretical model perfectly fits to the experimental STDs, the R-factor would be 0. However, the real experiment is obviously subjected to inaccuracies and errors that will always preclude the exact fitting even though the theoretical model is correct. For this reason, the R-factor constitutes an essential tool to evaluate the model quality. To note, the evaluating criterion is rather subjective and the researcher must always analyze where the similarities or discrepancies between both datasets are coming from. A good way to do this consists of graphically representing the relative STD profiles, and comparing them from a qualitative perspective [143-144]. Hence, disparate values can be individually analyzed to seek for reasonable explanations.

Deviations could not only be originated by experimental errors, but they could also arise from other factors related to docking errors, wrong ligand conformations or complex recognition events involving the contribution of several binding modes. In most of cases, the fittings via CORCEMA are firstly improved by optimizing the theoretical parameters included in calculations, approximating them as much as possible to the real ones. These variables include kinetic and thermodynamic parameters (K_d, k_{on}, k_{off}, correlation times…) and other magnitudes related to experimental conditions (magnetic field, protein and ligand concentrations, selective saturation…). Additionally, a refinement protocol can be run in parallel to reduce the R-factor even below 0.1 [145]. For this refinement, the same software has been designed to introduce other experimental restrictions like trNOEs, intra-ligand relaxation rates and torsion angles [146].

3.5. Transferred-NOESY: unveiling the bound conformation

The transferred NOESY (tr-NOESY) or exchange-transfer NOESY (et-NOESY) shares common features with the STD experiment. It is also based on the transference of magnetization though cross-relaxation mechanisms existing between spatially close nuclei, and then, it turns out to be particularly useful for receptor-ligand systems in which both counterparts largely differ in their

respective sizes. From a conceptual perspective, the experiment is fairly simple: a 2D-NOESY sequence with a mixing time optimized for the receptor, which is essentially a large protein in most of cases [130,147-148]. Thus, in contrast to the STD pulse sequence, the sample is not irradiated into a specific range of frequencies, but it is simply allowed to evolve along this mixing time in which through-space correlations develop as NOE crosspeaks. In this scenario, the cross relaxation rate of a given ligand is contributed by both mechanisms independently operating in the free and the bound state [149]:

$$< \sigma >= x_F \sigma_F + x_B \sigma_B \qquad (Eq. 17)$$

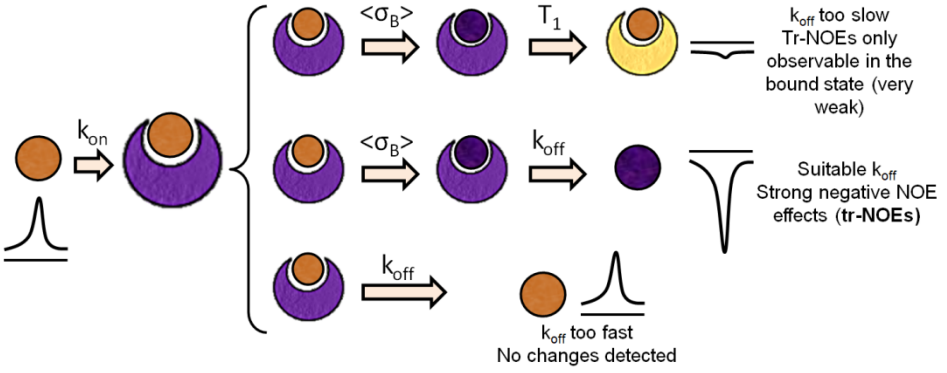


Figure 3.13. *The possible events taking place in a 2D transferred-NOESY, depending on the specific values of k_{off}, σ_B and T_1. The ideal situation for observing tr-NOEs is that represented in the middle, where the k_{off} is slow enough to firstly enable intermolecular cross relaxation (σ_B) and fast enough to allow the ligand to leave the receptor before complete relaxation (T_1), thus exhibiting the transferred NOE information in the free state.*

As previously exposed, the NOE mechanisms deeply depend on the correlation time (τ_C), which is simultaneously ligated to the molecular size. Consequently, the contribution of the free ligand term is considerably lower than that of the bound ligand, even approaching zero for biomolecules like oligosaccharides, whose molecular sizes are within the 500-1000 Da range. In the bound state, the ligand is experimenting a correlation time remarkably higher, derived from its attachment to a large slow-tumbling protein. As a result, the NOE crosspeaks gain intensity in accordance with the expectedly favored W_0 pathway. Moreover, derived from this mechanism, the intensity improvement is always accompanied with a sign switch from positive (or zero) to negative. In conclusion, this change on the sign of the ligand peaks serves as an unequivocal proof of ligand binding [150-152].

Similar to the STD, the tr-NOESY experiment becomes useful when applied to a system in the fast exchange regime (Figure 3.13). Under these conditions, the stronger NOE effects originated in the bound state are preserved when the ligand leaves the recognition site, giving rise to very sharp and intense signals upon longitudinal T_1 relaxation [153]. Then, the tr-NOESY spectrum is routinely performed in the presence of large protein/ligand ratios, typically ranging from 1:5 to 1:40, to achieve good signal-to-noise ratios. However, in the STD experiment the saturation is exclusively originated at the receptor, whereas herein all the ligand molecules subjected to the 2D-NOESY pulse sequence are susceptible to evolve through the NOE relaxation pathways. As a result, lower protein/ligand ratios than these used for STD are often preferred to minimize the contribution of the predominantly positive NOE arising from the large amount of free ligand molecules.

The aforementioned features underlie the key application of the tr-NOESY for interaction studies: the use of the transferred NOEs to deduce the ligand conformation in the bound state [154-156]. This application requires using peak integration to quantitatively determine inter-proton distances, and hence several warnings must be taken into account. As explained for the STD, the NOE intensities are affected by multiple factors whose influence may sometimes lead to a wrong description of the bound ligand conformation. Firstly, as mentioned above, the exchange regime is important as the NOE effect is attenuated for ligands that bind either very tightly ($K_d < 10^{-8}$ M) or very weakly ($K_d > 10^{-4}$ M) [149]. In the former case ($k_{off} << \sigma_B$), the NOE effects are notably reduced as the residence time inside the binding site is so long, where the tr-NOEs only become observable in the much less concentrated bound state and hence practically undetectable. Likewise, the latter case ($k_{off} >> \sigma_B$) leads to incomplete transference of the protein-edited NOE as complex dissociation is too fast. If the k_{off} is approximately known, the efficiency of the NOE experiment may be valued by comparison with the cross relaxation rate (σ_B), which can be roughly estimated as a function of the molecular size through the following empirical relationship (ignoring the σ_F term) [153]:

$$\boldsymbol{\sigma_{ij}}\ (s^{-1}) \approx\ -\mathbf{24} \cdot \frac{\mathbf{MW}\ (\mathrm{kDa})}{\mathbf{r_{ij}^{6}}\ (\text{Å})} \qquad (\mathrm{Eq.}\ 18)$$

Obviously, the mixing time (t_{mix}) must be also carefully chosen to allow cross relaxation mechanisms to efficiently progress through the ligand structure. Common values are usually set below 200 ms. Due to the difficulty to correctly

select a suitable t_{mix}, the NOE is often overrated because of the effects of spin diffusion [157-158]. Briefly, spin diffusion is the transference of magnetization between two relatively separated nuclei through a third intermediate atom that is located near to both nuclei. As a result, the NOE crosspeak intensity is virtually enhanced (or reduced) and its integral value is consequently misestimated, whereby often concluding that those protons lie closer to each other than they really are. The negative effects of the spin diffusion have been repeatedly reported [159-160], and they can arise both from intramolecular and intermolecular dipolar couplings. The predominant effect (intra/inter) will depend on how the intermediate nucleus is spatially arranged respect to the other two (Figure 3.14). Graphically, the alternative ROESY sequence is the most suitable method to detect spin diffusion peaks in a 2D spectrum, since they appear as signals with opposite sign respect to the real NOEs [161-162].

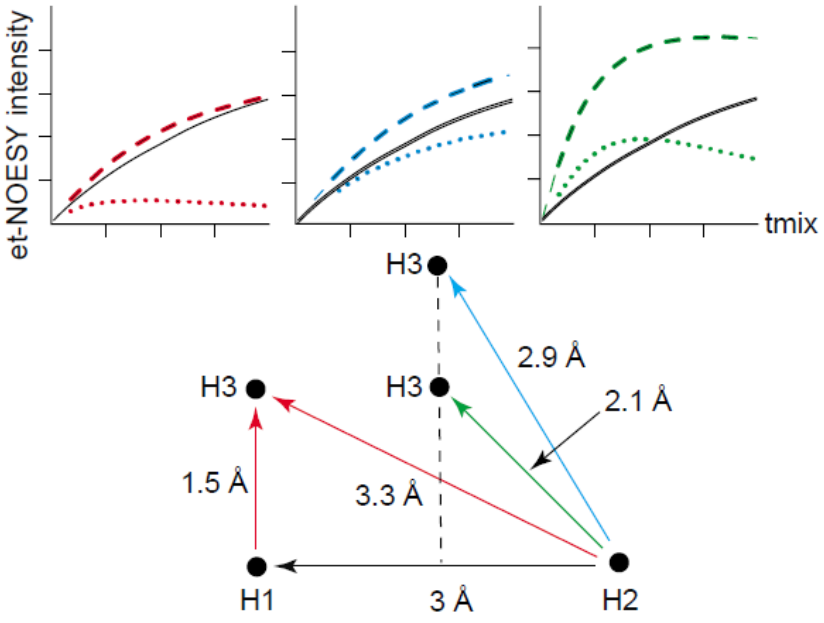


Figure 3.14. *Effect of the spin diffusion mechanism on the evolution of the NOE, depending on different geometrical factors. On the top, the NOE build-up curves considering different locations of the intermediate nucleus (H3), represented on the bottom (red/blue/green). The black curves represent the ideal case of an isolated two-spin system. The colored dashed and dotted curves respectively represent the NOE evolution in the presence of intermolecular H3 or intramolecular H3. Adapted from [159].*

From a strict point of view, the quantitative determination of H-H distances is usually made by applying the *initial slopes* strategy again. To completely erase the undesired spin diffusion effects, the NOE magnitudes for the different H-H

pairs are related to the initial slopes extracted from the exponential fitting of the NOE integral versus the mixing time. Of course, the time cost is higher as several 2D-NOESY spectra at different t_{mix} must be recorded. Apart from spin diffusion, other problems may still arise from the protein background and the non-specific binding phenomena at high protein-ligand ratios [163]. Even using a large ligand excess, the intramolecular NOEs of the protein may hinder the precise determination of peak integrals, introducing moderate mistakes in the estimated distances. In this regard, the application of T_2 or $T_{1\rho}$ filters to suppress the protein signals is highly recommended if quantitative analyses are going to be performed. Non-specific binding is a minor problem that does not seriously affect the determination of bound conformations in most of cases. There are several options to prove whether a recognition process is specific, although the most straightforward method consists of adding another specific ligand to check by competition the NOE intensity decay in the compound of interest.

3.6. ^{19}F-based T_2 relaxation

Within the aforementioned ligand-based approaches, nuclei relaxation is likewise applied to the study of binding events involving two entities with different sizes [164-165]. The mathematical expressions that provide the expected relaxation rates (R_1 and R_2) for a given system are rather complex and include several parameters, including dipolar and scalar couplings, chemical shift anisotropy (CSA) and chemical exchange [166-167]. However, an important variable shared by most of these factors is the correlation time (τ_C), also understood as the time that a given molecule needs to complete a turn around its own rotation axis (Figure 3.15). This time usually ranges from 0.1 ns for small molecules to several tens of nanoseconds for typical macromolecules (e.g. 41 ns for BSA [168]). Therefore, observable changes in the line broadenings of a small molecule after adding a large receptor can be attributed to a binding event. The experimental methodology is essentially based on recording two experiments with the compound of interest, in the absence and in the presence of the receptor [169]. Indeed, the relative simplicity of this experimental setup has made this technique a common choice for accomplishing high throughput screening (HTS) [170-172]. Nevertheless, the measurement of relaxation rates requires a clean observation of the attenuated signals, something often difficult in ^{1}H-NMR, especially when applied to the screening of mixtures containing many compounds [173]. A described solution consist of subtracting one spectrum (with receptor) from the other one (free compounds), obtaining a third

spectrum similar to the STD experiment, that erases the signals from the non-interacting compounds, exclusively leaving those from the bound ones [174]. Other approaches have relied on including T_2 filters in 2D experiments like COSY or TOCSY, thus alleviating the signal crowding for better observing line widths [175]. Apart, a popular method broadly used to improve the sensitivity of the experiment consists of incorporating paramagnetic tags (e.g. TEMPO) to enhance the transversal relaxation rates [176], either attached to the macromolecule [177] or to another ligand already known [178]. The latter approach is generally employed to detect binders at secondary sites, placing the tagged ligand at the primary binding site.

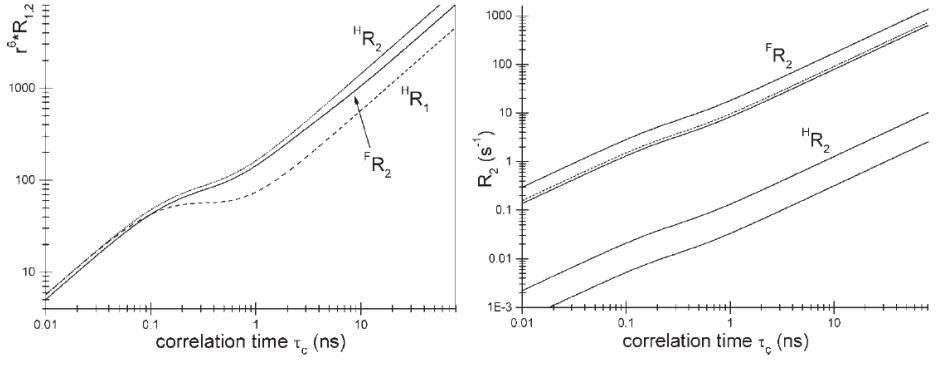


Figure 3.15. *The evolution of the transversal (R_2) and longitudinal (R_1) relaxation rates for ^{19}F and ^{1}H as a function of the correlation time, τ_C. On the left, the dependence originated by dipolar couplings, and on the right, the contribution arising from the chemical shift anisotropy (CSA). Adapted from [167].*

Aside from these options, research has progressively pointed towards the utilization of the ^{19}F nucleus instead of the proton. The ^{19}F isotope is very similar to ^{1}H in terms of natural abundance and gyromagnetic ratio, which makes it as good NMR probe as the proton. Moreover, it offers numerous advantages from a practical point of view [179], and specifically for relaxation studies:

- Whereas protons are intrinsically present in every (bio)organic molecule, the biological compounds are completely devoid of ^{19}F. Then, this nucleus is always intentionally introduced in the analyzed ligands by chemical synthesis. Since only one or few fluorine atoms are generally introduced per molecule, the resulting spectra are considerably cleaner, displaying one signal per compound and eliminating the background arising from the receptor, the solvent and other components [53,180].

- The fluorine atom is remarkably more sensitive to its electronic environment, whereby NMR signals are typically spread over several tens of ppms. Together with the previous point, these features allow readily working with large mixtures of fluorinated compounds avoiding signal overlapping problems.
- As a result from the chemical sensitivity mentioned in the latter point, the ^{19}F nucleus becomes an excellent probe in order to measure transverse relaxation rates ($R_2 = 1/T_2$) for two reasons: firstly, the contribution of the chemical shift anisotropy (CSA) of the ^{19}F to its global R_2 leads to notable differences in the line widths between the free and the bound states [181]. Moreover, this extreme sensitivity is accentuated as the chemical environment is more perturbed. Then, the detection limit of these experiments can be expectedly improved at higher fields [182]. Secondly, the chemical exchange between both possible states (free/bond) when a receptor is present also makes T_2 faster [183]. The mathematical expression that determines the contribution of the chemical exchange to R_2 is known as the Swift-Connick equation (Eq. 19, [184]). Usually, since [L] >>> [P] in ligand-based methods, the fraction of free ligand (x_F) is considered as 1:

$$R_{2,exc} = x_F R_{2,free} + \frac{x_F x_B}{\tau_{res}} \left[\frac{R_{2B}\left(R_{2B} + \frac{1}{\tau_{res}}\right) + 4\pi^2(\delta_F - \delta_B)^2}{R_{2B}\left(R_{2B} + \frac{1}{\tau_{res}}\right)^2 + 4\pi^2(\delta_F - \delta_B)^2} \right] \quad (Eq.19)$$

  Where x_F/δ_F and x_B/δ_B are the molar fractions and chemical shifts of the free and bound forms, respectively, R_{2B} is the transverse relaxation rate of the ligand in the bound state, and τ_{res} is the residence time in the bound state, also interpreted as $1/k_{off}$.

The dependence of the exchange contribution on the residence time (or $1/k_{off}$), taking into account the other parameters (R_{2B}, $\delta_B - \delta_F$), has been evaluated in detail many times, especially by C. Dalvit and co-workers [167,182], and is crucial to determine the range of applicability of ^{19}F-T_2 experiments in the analysis of molecular interactions (Figure 3.16). These studies have highlighted that the chemical exchange is certainly one of the most relevant effects regulating the ligand relaxation rates for most of the protein-ligand systems usually investigated. In fact, when the difference in chemical shifts ($\delta_B - \delta_F$) is large enough, the exchange contribution is notorious even at very high protein/ligand ratios (low x_B). This implies a double advantage: on one hand, for ligands with moderate affinities, it is feasible to

use really low amounts of receptor, even reaching pM concentrations [185]. On the other hand, the upper affinity limit becomes higher since weak binders (K_d = 10^{-3}-10^{-2} M) can be also detected as long as their k_{off} values are of the same order as $|\delta_B - \delta_F|$. In addition, as deduced from the Swift-Connick expression, the influence of the chemical exchange does not depend on the protein correlation time, thereby enabling its effective application to small receptors as well [182].

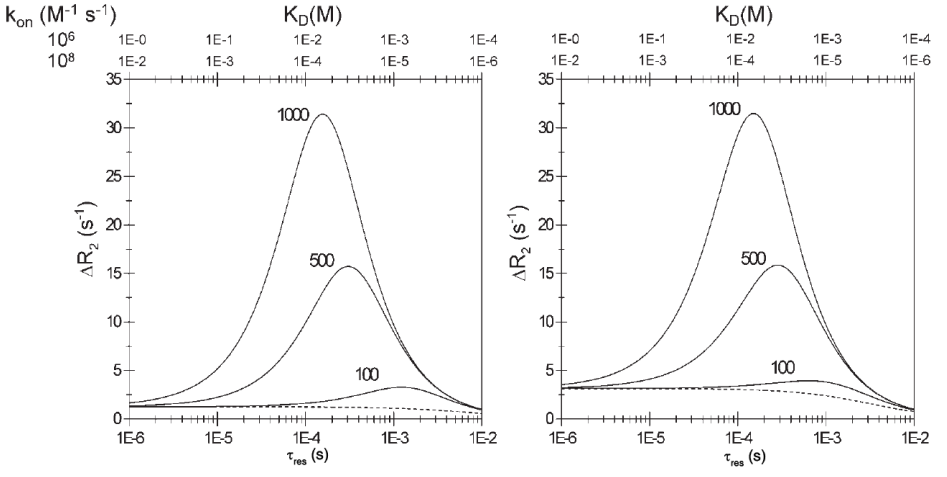


Figure 3.16. *Contribution of the chemical exchange to the transversal relaxation rate (R_2) as predicted by the Swift-Connick equation, depending on τ_{res} (the inverse of k_{off}). The two graphs are calculated for two different R_{2B}, and each graph displays the variation predicted for three different δ_B - δ_F, expressed in Hz (100, 500 and 1000 Hz).The upper scale shows the expected value of the dissociation constant in two different cases, for k_{on} = 10^6 $(Ms)^{-1}$ and k_{on} = 10^8 $(Ms)^{-1}$. Adapted from [167].*

Experimentally, the relaxation rates are normally determined by means of two experiments: a routine ^{1}H-1D in which the intensity of the fluorine signal is measured after a 90º pulse and an additional R_2-filtered ^{1}H-1D in which the intensity of the fluorine signal is modulated after the initial 90º pulse by a spin-echo scheme, based on applying a train of 180º pulses separated by a delay "2τ" [186-187]. As a result, the latter experiment is merely a ^{1}H-1D with an additional element that makes the fluorine signal to get weaker proportionally to its specific T_2. This particular spin-echo scheme, as depicted in Figure 3.17A, is commonly known as the Carr-Purcell-Meibom-Gill (CPMG) pulse sequence [188-189]. The relationship between the observed T_2 ($T_{2,obs}$) and the signal attenuation is established as follows:

$$\mathbf{I = I_0 exp\left(-\frac{n2\tau}{T_{2,obs}}\right)} \qquad (Eq.\,20)$$

Where I_0 is the signal intensity in the normal ^{1}H-1D spectrum, and "n" is the number of times that the 2τ + 180º interval is repeated (number of echoes). This is the simplest version of such an equation, since the complete expression includes another exponential term that adds the contribution of translational diffusion ("D") and the magnetic field inhomogeneities (ΔB_0) to the signal decay [182]. However, these contributions only become really significant when the 2τ period is too long, and hence, the value of "τ" should be reasonably chosen to work without considering that term. Analogously, if the 2τ interval is too short, the evolution of the chemical shift is not suitably driven and the effect of chemical exchange is partially lost [190], what may deeply affect to the experiment sensibility. Generally, the number of echoes ("n") is adjusted through a previous evaluation of the $T_{2,free}$ and the T_2 observed in the presence of the protein ($T_{2,obs}$). By recording several CPMG spectra at different values of "n", both decay curves are jointly represented and compared (Figure 3.17B). The best values ranges within the intermediate zone where it is high enough to appreciate how the ligand signal has been differently affected in the presence and absence of the receptor [191]. At longer or shorter "n" values, both signals (free/bound) display similar percentages of attenuation and comparisons may become unrealistic. Of course, T_2 can be roughly estimated by directly applying Eq. 20 with a single CPMG experiment (with "n" and "τ" well adjusted), but the most accurate procedure by far consists of fitting the data from the aforementioned decay curves to Eq. 20. This is undoubtedly the best method not only to calculate transverse relaxation rates but also to evaluate whether there is specific binding or not [191]. In this regard, the acquisition of an only CPMG experiment is often reserved for screening purposes, namely detecting potential binders (hits) for a given receptor through the direct observation of line width changes in a fluorinated probe (spy molecule) [192-193], for which the experimental setup has been previously optimized as detailed above.

Competition experiments using a fluorinated compound as "spy molecule" have been repeatedly employed in order to indirectly determine binding constants of other specific binders of the same receptor [194]. For this purpose, the K_d of the spy molecule (K_S) must be known a priori. With the dissociation constant, it is possible to readily determine x_B at a given $[S]_0$ through Eq. 10, as already explained. Then, different values of x_B are determined at different $[P]_0$, and correlated with the corresponding signal decays experimentally measured in each case. The resulting exponential fit is subsequently used to interpolate new values of x_B when the tested compound (competitor, C) is added to the mixture of receptor (P) and ^{19}F-probe (S). With

these data, the K_S of the probe is recalculated as an apparent K_S (K_S^{app}) for each concentration of "C" used, and both variables are finally adjusted to Eq. 21, which yields the dissociation constant of the tested molecule, K_C.

$$K_C = \frac{[C]K_S}{K_S^{app} - K_S} \qquad (Eq. 21)$$

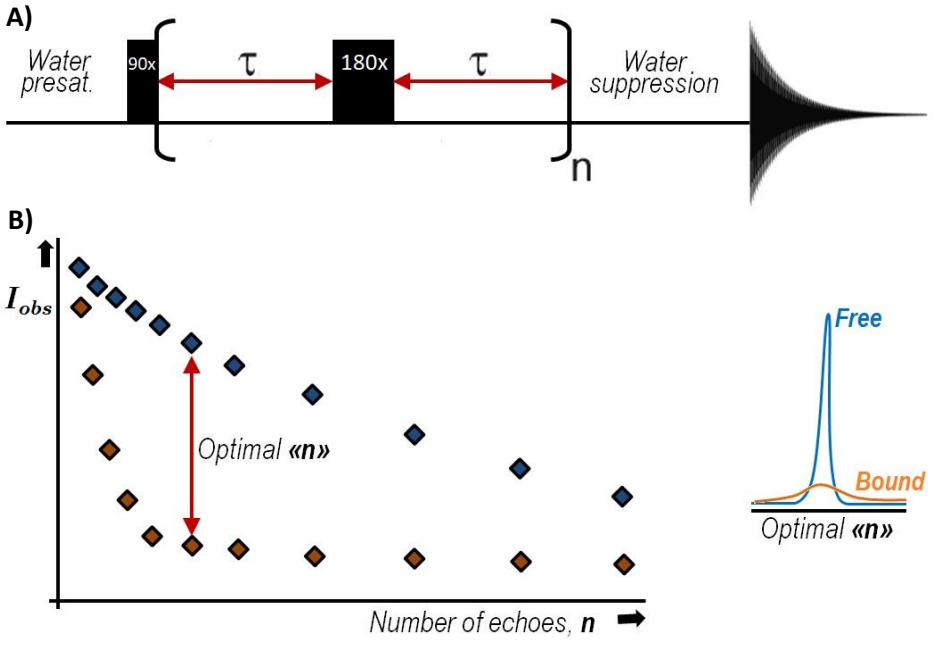


Figure 3.17. *A) The CPMG pulse sequence which provides an R_2-filtered FID. After the hard pulse (90x), the 2τ+180x segment (echo) is repeated "n" times. "τ" and "n" are parameters of the Eq. 20. B) The signal decays observed at different "n" values for a relatively small compound in the absence (blue) and the presence (orange) of a much larger receptor. The optimal value of "n" to differentiate both states (free/bound) is graphically determined.*

3.7. 2D-DOSY for estimating the protein size

The random translational motion that every molecule experiences in solution due to its internal kinetic energy is referred to as self-diffusion [195]. The parameter used to describe this phenomenon is called the diffusion coefficient ("D"), which quantifies the translational speed of a molecule.

Self-diffusion is influenced by the physical properties of both the proper molecule (shape, size, charge) and the surrounding milieu (temperature, viscosity). Analogously to T_2-filtered experiments, these physical properties can be exploited by their influence in the relaxation of active nuclei to turn

diffusion into an observable NMR parameter [196-197]. Specifically, diffusion coefficients are routinely evaluated by pulsed-field gradient (PFG) sequences [198]. Broadly speaking, these experiments are based on the application of a gradient pulse that generates a variable magnetic field along the NMR tube (the "z" axis). Under the influence of this magnetic gradient, each molecule experiences a different Larmor frequency which depends on its specific spatial position. In other words, the phase of the measured spins turns out to be spatially encoded at this point (Figure 3.18A). After a certain delay (Δ, diffusion time), an echo pulse (180º) is applied to remove the contribution of the chemical shifts and then, another gradient pulse removes the effects of the first pulse by refocusing all the spins again. In the most ideal case, if all the molecules would remain static at their initial positions, the second gradient pulse would allow to fully recovering the initial signal. However, due to Brownian translational motion, the molecules have already moved randomly away from their positions, whereby the signal cannot be well refocused and it is inevitably attenuated due to self-diffusion. The higher the diffusion coefficient is, the faster the molecular displacement is and consequently the weaker the NMR signals will be. The peak attenuation can be expressed as an exponential function that essentially depends on the diffusion coefficient (D), the diffusion time (Δ, delay for allowing the spatially-driven encoding) and other parameters related to the gradient pulse (strength, "g", and duration, "δ") and the observed nucleus (γ):

$$\mathrm{I} = \mathrm{I}_0 \exp\left[-\mathrm{D}(\gamma \mathrm{g}\delta)^2\left(\Delta - \frac{\delta}{3}\right)\right] \qquad (Eq.\,22)$$

In the practice, this sequence is carried out several times by varying one of the parameters related to the PFG scheme (usually, the gradient pulse strength, "g"), giving rise to a series of ^{1}H-1D NMR spectra in which the signals from each compound have been differently affected by their diffusion coefficients. Subsequent data processing, based on Eq. 22, leads to a pseudo-2D spectrum in which the direct dimension shows the normal proton resonances, while the second dimension separates these resonances according to the diffusion coefficient, commonly represented in logarithmic scale. This experiment is better known as Diffusion Ordered SpectroscopY (DOSY) (Figure 3.19) [199-200].

The standard PGF scheme was firstly developed by Stejskal and Tanner [201] (Figure 3.18B), and was subsequently improved with modifications aimed at suppressing undesired artifacts arising from signal modulation caused by spin-spin couplings [202-203] and eddy currents (electric current loops) originated

by the presence of a variable B_0 [204]. More problematic than eddy currents are convection currents created by temperature gradients [205]. In both cases, the diffusion coefficients are overestimated since these currents increase the molecular motions throughout the NMR tube. In this regard, several practical solutions have permitted to easily compensate temperature gradients in most of cases, such as applying a faster gas flow, using solvents with higher viscosities or spinning the tube at certain speeds [206].

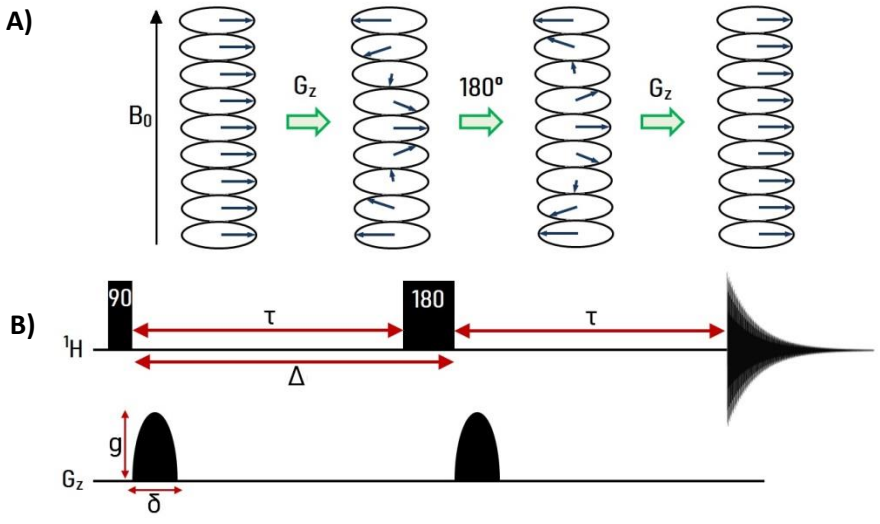


Figure 3.18. *A) Phase encoding based on the position of every nucleus along the variable B_0, generated by a gradient pulse G_Z. After the spin-echo (180º), the magnetization is refocused through an identical G_Z. B) The Stejskal and Tanner pulse sequence. Parameters: τ, delay between pulses; Δ, diffusion time; δ, gradient pulse duration; g, gradient pulse strength. Adapted from [201].*

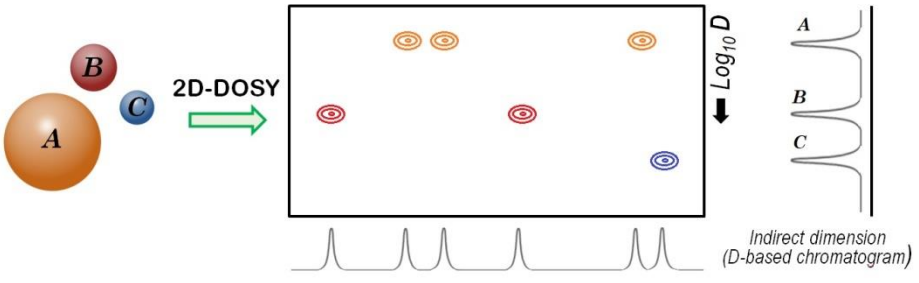


Figure 3.19. *The 2D-DOSY experiment, in which the normal ^{1}H-1D containing signals from three different compounds is expanded along a secondary dimension, thereby separating every signal from each compound according to its diffusion-driven attenuation. The vertical projection of this experiment (right) will display the three compounds independently separated and ordered by their respective diffusion coefficients (and hence, indirectly by their sizes).*

As stated above, diffusion coefficients can be related to certain physical properties. Under the same solution conditions, two or more compounds can be distinguished by their diffusion coefficients, which are closely correlated with their molecular sizes through the Stokes-Einstein equation (Eq. 23). Hence, 2D-DOSY NMR is applicable to the determination of molecular sizes [207-208] and indeed it is usually considered as a non-invasive *chromatography* [209], since the indirect dimension can be interpreted as a size-ordered *spin chromatogram* (Figure 3.19). Noteworthy, the separation of the observed nuclei in the indirect dimension is purely virtual, and those peaks from different compounds resonating at very close or identical chemical shifts will appear as a "mixed" peak at an intermediate value of the diffusion coefficient contributed by both compounds [210]. For these cases, if the analyzed molecules cannot be well recognized, the best practical solution is to extend the information to additional dimensions, thereby alleviating signal overlap. As examples, the DOSY methodology has been successfully combined with COSY (3D-DOSY-COSY) [211], TOCSY (3D-DOSY-TOCSY) [212] and even HSQC schemes (3D-DOSY-HSQC) [213]. Another potential solution, although more complex, would be a multiexponential fit of the recorded data.

$$D = \frac{k_B T}{6\pi\eta r_H} \qquad (Eq. 23)$$

In the Stokes-Einstein equation, the hydrodynamic radius (r_H) gives an estimation of the molecular size under the assumption of a spherical shape and supposing that the solvent is a continuous medium. Several other authors have updated this equation with additional terms that account for the size of the solvent molecules [214-215] and the non-spherical shape of the studied molecules [216-217]. In any case, the direct determination of this term once the diffusion coefficient is determined by NMR is rather unpractical, as it requires the previous determination of the solution viscosity (η). To avoid the introduction of large errors derived from additional calculations or approximations, a more straightforward method consists of directly comparing the diffusion coefficients of the unknown molecule and other reference molecule, whose size features (r_H, MW) have been already determined or reported [218]. As has been done in this work, such an approach, if applied to macromolecules, would imply adding at least another reference protein whose size is in the same range as the studied system (e.g. a commercial protein like BSA) and it has to be measured in the same experimental conditions as the problem sample. Assuming certain approximations, the Stokes-Einstein equation can be adapted to be expressed

as a function of the molecular weight (MW) instead of r_H, and thereby the unknown molecular weight can be directly estimated as follows:

$$D = \frac{k_B T}{6\pi\eta\left[\frac{3 \cdot MW}{4\pi N\rho}\right]^{\frac{1}{3}}} \quad ; \quad \frac{D}{D_{ref}} = \frac{\frac{k_B T}{6\pi\eta\left[\frac{3 \cdot MW}{4\pi N\rho}\right]^{\frac{1}{3}}}}{\frac{k_B T}{6\pi\eta\left[\frac{3 \cdot MW_{ref}}{4\pi N\rho}\right]^{\frac{1}{3}}}} \rightarrow$$

$$\frac{\mathbf{D}}{\mathbf{D_{ref}}} = \sqrt[3]{\frac{\mathbf{MW_{ref}}}{\mathbf{MW}}} \qquad (Eq. 24)$$

Where D and D_{ref} are estimated by 2D-DOSY, and MW_{ref} is the molecular weight of the reference protein. If several reference compounds are available (e.g. a small library of commercial proteins), a better alternative consist of fitting the data to an approximate linear regression that would correlate the $\log_{10}$ D with the $\log_{10}$ MW, and then interpolate the unknown molecular weight once the experimental diffusion coefficient has been determined [219-220].

It is worth mentioning that the 2D-DOSY also works as a ligand-based method to detect binding, as it is likewise based on observing a size-dependent NMR parameter [221]. However, its application is fairly limited to systems in which the exchange regime is predominantly fast, and thereby the ligand will expectedly exhibit a much lower diffusion coefficient in the presence of a large receptor [222-223]. For slow and intermediate exchange, if the ligand is very small, detection of binding become problematic as the free and bound states may appear as separated signals, and therefore, the signal of the free ligand could be interpreted as absence of binding and the signal of the bound ligand, as the free receptor (approx. the same MW).

3.8. References

1. J. Etxebarria, J. Calvo, M. Martin-Lomas and N. C. Reichardt. Lectin-array blotting: profiling protein glycosylation in complex mixtures. *ACS Chem. Biol.,* (2012), **7**, 1729-1737.
2. O. Blixt, S. Head, T. Mondala, C. Scanlan, M. E. Huflejt, R. Alvarez, M. C. Bryan, F. Fazio, D. Calarese, J. Stevens, N. Razi, D. J. Stevens, J. J. Skehel, I. van Die, D. R. Burton, I. A. Wilson, R. Cummings, N. Bovin, C. H. Wong and J. C. Paulson.

Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl. Acad. Sci. U S A,* (2004), **101**, 17033-17038.

3. P. H. Liang, C. Y. Wu, W. A. Greenberg and C. H. Wong. Glycan arrays: biological and medical applications. *Curr. Opin. Chem. Biol.,* (2008), **12**, 86-92.
4. J. R. Crowther. Basic Principles of ELISA. In *ELISA: Theory and Practice*, J. R. Crowther. Humana Press. Springer Science+Business Media (Totowa, New Jersey), 1995, **ch. 2**.
5. Y. Koshi, E. Nakata, H. Yamane and I. Hamachi. A fluorescent lectin array using supramolecular hydrogel for simple detection and pattern profiling for various glycoconjugates. *J. Am. Chem. Soc.,* (2006), **128**, 10413-10422.
6. Y. Ji and R. J. Woods. Quantifying weak glycan-protein interactions using a BioLayer Interferometry competition assay: Applications to ECL lectin and X-31 influenza hemagglutinin. *Adv. Exp. Med. Biol.,* (2018), **1104**, 259-273.
7. A. Gruden-Movsesijan and Lj. S. Milosavljevic. The involvement of the macrophage mannose receptor in the innate immune response to infection with parasite *Trichinella spiralis*. *Vet. Immunol. Immunopathol.,* (2006), **109**, 57-67.
8. a) E. Duverger, N. Lamerant-Fayel, N. Frison and M. Monsigny. Surface plasmon resonance: methods and protocols. In *Methods in Molecular Biology*, eds. N.J. De Mol and M.J.E. Fischer. Springer (Utrecht), (2010), **627**, 157–178. b) A. Penezic, G. Deokar, D. Vignaud, E. Pichonat, H. Happy, P. Subramanian, B. Gasparovic, R. Boukherroub and S. Szunerits. Carbohydrate-lectin interaction on graphene-coated Surface Plasmon Resonance (SPR) interfaces. *Plasmonics*, (2014), **9**, 677-683.
9. a) T. K. Dam, M. L. Talaga, N. Fan and C. F. Brewer. Measuring multivalent binding interactions by Isothermal Titration Calorimetry. *Methods Enzymol.,* (2016), **567**, 71-95. b) B. N. Murthy, S. Sinha, A. Surolia, S. S. Indi and N. Jayaraman. SPR and ITC determination of the kinetics and the thermodynamics of bivalent versus monovalent sugar ligand-lectin interactions. *Glycoconj. J.,* (2008), **25**, 313-321. c) Z. Wang, G. Chen, J. Lu, L. Hong and T. Ngai. Investigation of the factors affecting the carbohydrate–lectin interaction by ITC and QCM-D. *Colloid. Polym. Sci.,* (2014), **292**, 391-398.
10. J. Agirre. Strategies for carbohydrate model building, refinement and validation. *Acta Crystallogr. D. Struct. Biol.,* (2017), **73**, 171–186.
11. a) B. S. Blaum, U. Neu, T. Peters and T. Stehlea. Spin ballet for sweet encounters: saturation-transfer difference NMR and X-ray crystallography complement each other in the elucidation of protein–glycan interactions. *Acta Crystallogr. F Struct. Biol. Commun.,* (2018), **74**, 451–462. b) V. Roldos, F. J. Canada and J. Jimenez-Barbero. Carbohydrate-protein interactions: a 3D view by NMR. *Chembiochem.,* (2011), **12**, 990–1005.
12. R. Wild, J. Kowal, J. Eyring, E. M. Ngwa, M. Aebi and K. P. Locher. Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science,* (2018), **359**, 545–550.

13. a) V. Chaptal, S. Kwon, M. R. Sawaya, L. Guan, H. R. Kaback and J. Abramson. Crystal structure of lactose permease in complex with an affinity inactivator yields unique insight into sugar recognition. *Proc. Natl. Acad. Sci.,* (2011), **108**, 9361-9366. b) H. Ideo, T. Matsuzaka, T. Nonaka, A. Seko and K. Yamashita. Galectin-8-N-domain recognition mechanism for sialylated and sulfated glycans. *J. Biol. Chem.,* (2011), **286**, 11346-11355.
14. D. Sirohi, Z. Chen, L. Sun, T. Klose, T. C. Pierson, M. G. Rossmannand and R. J. Kuhn. The 3.8 Å resolution cryo-EM structure of Zika virus. *Science,* (2016), **352**, 467–470.
15. a) J. Agirre, G. J. Davies, K. S. Wilson and K. D. Cowtan. Erratum. Carbohydrate anomalies in the PDB. *Nat. Chem. Biol.,* (2015), **11**, 303. b) J. Agirre, G. J. Davies, K. S. Wilson and K. D. Cowtan. Carbohydrate structure: the rocky road to automation. *Curr. Opin. Struct. Biol.,* (2017), **44**, 39-47.
16. J. R. Helliwell. Protein crystal perfection and its application. *Acta Crystallogr., Sect. D: Biol. Crystallogr.,* (2005), **61**, 793–798.
17. J. Dubochet and E. Knapek. Ups and downs in early electron cryo-microscopy. *PLoS Biol.,* (2018), **16**, e2005550.
18. M. Adrian, J. Dubochet, J. Lepault and A. W. McDowall. Cryo-electron microscopy of viruses. *Nature,* (1984), **308**, 32–36.
19. S. Berger and S. Braun. *200 and More NMR Experiments*. WILEY-VCH Verlag GmbH & Co. KGaA (Weinheim), 1st edn, 2004.
20. J. L. Asensio, H. C. Siebert, C. W. von Der Lieth, J. Laynez, M. Bruix, U. M. Soedjanaamadja, J. J. Beintema, F. J. Cañada, H. J. Gabius and J. Jiménez-Barbero. NMR investigations of protein-carbohydrate interactions: studies on the relevance of Trp/Tyr variations in lectin binding sites as deduced from titration microcalorimetry and NMR studies on hevein domains. Determination of the NMR structure of the complex between pseudohevein and N,N',N''-triacetylchitotriose. *Proteins,* (2000), **40**, 218-326.
21. M. F. García-Mayoral, L. García-Ortega, M. P. Lillo, J. Santoro, A. Martínez del Pozo, J. G. Gavilanes, M. Rico and M. Bruix. NMR structure of the noncytotoxic-sarcin mutant(7-22): The importance of the native conformationof peripheral loops for activity. *Protein Sci.,* (2004), **13**, 1000–1011.
22. S. W. Homans. NMR spectroscopy tools for structure-aided drug design. *Angew. Chem. Int. Ed. Engl.,* (2004), **43**, 290-300.
23. J. Vaynberg and J. Qin. Weak protein-protein interactions as probed by NMR spectroscopy, *Trends Biotechnol.,* (2006), **24**, 22–27.
24. A. Mallagaray, C. Rademacher, F. Parra, G. Hansman and T. Peters. Saturation transfer difference nuclear magnetic resonance titrations reveal complex multistepbinding of L-fucose to norovirus particles. *Glycobiology,* (2017), **27**, 80–86.
25. K. Scheffler. B. Ernst, A. Katopodis, J. L. Magnani, W. Wang, R. Weiseinann and T. Peters. Determination of the bioactive conformation of the carbohydrate

ligand in the E-Selectin/Sialyl LewisY complex. *Angew. Chem. Int. Ed. Engl.,* (1995), **34**, 1841-1844.

26. A. Canales, I. Boos, L. Perkams, L. Karst, T. Luber, T. Karagiannis, G. Domínguez, F. J. Cañada, J. Pérez-Castells, D. Häussinger, C. Unverzagt and J. Jiménez-Barbero. Breaking the limits in analyzing carbohydrate recognition by NMR spectroscopy: Resolving branch-selective interaction of a tetra-antennary N-glycan with lectins. *Angew. Chem. Int. Ed. Engl.,* (2017), **56**, 14987-14991.
27. K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübner and D. Kern. Intrinsic motions along an enzymatic reaction trajectory. *Nature,* (2007), **450**, 838-844.
28. B. Fiege, M. Leuthold, F. Parra, K. P. Dalton, P. J. Meloncelli, T. L. Lowary and T. Peters. Epitope mapping of histo blood group antigens bound to norovirus VLPs using STD NMR experiments reveals fine details of molecular recognition. *Glycoconj. J.,* (2017), **34**, 679–689.
29. M. Rinnbauer, B. Ernst, B. Wagner, J. Magnani, A. J. Benie and T. Peters. Epitope mapping of sialyl LewisX bound to E-selectin using saturation transfer difference NMR experiments. *Glycobiology,* (2003), **13**, 435-443.
30. B. S. Blaum, U. Neu, T. Peters and T. Stehlea. Spin ballet for sweet encounters: saturation-transfer difference NMR and X-ray crystallography complement each other in the elucidation of protein−glycan interactions. *Acta Crystallogr., Sect. F: Struct. Biol. Commun.,* (2018), **74**, 451−462.
31. T. D. W. Claridge. *High resolution NMR techniques in Organic Chemistry*. Elsevier Science (Amsterdam), 3rd edn, 2016.
32. N. E. Jacobsen. *NMR spectroscopy explained*. John Wiley & Sons (Hoboken, New Jersey), 1st edn, 2007.
33. J. Keeler. *Understanding NMR spectroscopy*. John Wiley & Sons (Hoboken, New Jersey), 2nd edn, 2010.
34. L. M. Harwood and T. D. W. Claridge. *Introduction to Organic Spectroscopy*. Oxford University Press (Oxford), 1st edn, 1997.
35. A. E. Ferenz and G. Wagner. NMR spectroscopy: a multifaced approach to macromolecular structure. *Quarter. Rev. Biophys.,* (2000), **33**, 211-240.
36. D. A. Case, H. J. Dyson and P. E. Wright. Use of chemical shifts and coupling constants in NMR structural studies of peptides and proteins. *Methods Enzymol.,* (1994), **239**, 393-416.
37. K. Wüthrich. *NMR of proteins and nucleic acids*. John Wiley & Sons (New York), 1st edn, 1986, pp 117-129.
38. M. Allegrozzi, I. Bertini, M. B. L. Janik, Y. M. Lee, G. H. Lin and C. Luchinat. Lanthanide-induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 Å from the metal ion. *J. Am. Chem. Soc.,* (2000), **122**, 4154-4161.

39. C. A. G. Haasnoot. Conformational analysis of six-membered rings in solution: ring puckering coordinates derived from vicinal NMR proton-proton coupling constants. *J. Am. Chem. Soc.,* (1993), **115**, 1460-1468.
40. C. Altona and C. A. G. Haasnoot. Prediction of anti and gauche vicinal proton-proton coupling constants in carbohydrates: A simple additivity rule for pyranose rings. *Magn. Res. Chem.,* (1980), **13**, 417-429.
41. T. Church, Ian Carmichael and A. S. Serianni. Two-bond 13C-13C spin-coupling constants in carbohydrates: effect of structure on coupling magnitude and sign. *Carbohydr. Res.,* (1996), **280**, 177-186.
42. G. W. Vuister, M. Tessari, Y. Karimi-Nejad and B. Whitehead. Modern techniques in protein NMR. *Biol. Magn. Reson.,* (1999), **16**, 195-257.
43. F. Cordier and S. Grzesiek. Direct observation of hydrogen bonds in proteins by interresidue $^{3h}J_{NC'}$ scalar couplings. *J. Am. Chem. Soc.,* (1999), **121**, 1601–1602.
44. M. Karplus. Contact electron-spin coupling of nuclear magnetic moments. *J. Phys. Chem.,* (1959), **30**, 11-15.
45. P. Schmieder and H. Kessler. Determination of the phi-angle in a peptide backbone by NMR-spectroscopy with a combination of homonuclear and heteronuclear coupling-constants. *Biopolymers,* (1992), **32**, 435-440.
46. N. Tjandra and A. Bax. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science,* (1997), **278**, 1111-1114.
47. D. Neuhaus. Nuclear Overhauser Effect. In *Encyclopedia of Magnetic Resonance (eMagRes)*. John Wiley & Sons (Hoboken, New Jersey). 2011.
48. D. D. Traficante. Relaxation. Can T2, be longer than T1? *Concepts Magn. Reson.,* (1991), **3**, 171–177.
49. M. H. Levitt. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. John Wiley & Sons (New York), 2nd edn, 2008, section 11.9.2.
50. O. Zerbe and S. Jurt. *Applied NMR Spectroscopy for Chemists and Life Scentists*. Wiley-VCH Verlag GmbH & Co. KGaA (Weinheim), 1st edn, 2014.
51. B. J. Stockman and C. Dalvit. NMR screening techniques in drug discovery and drug design. *Prog. Nucl. Magn. Reson. Spectrosc.,* (2002), **41**, 187–231.
52. V. N. Sivanandam, G. Bernardo-Seisdedos and O. Millet. Receptor-based NMR techniques in drug discovery. In *Biophysical Techniques in Drug Discovery*, eds. A. Canales, D. Thurston, A. Martinez, A. Scott, M. J. Eggen, J. F. Espinosa, O. Millet, J. Bouckaert, M. Molina, L. Mafra, S. Pollack, S. Ortega, O. Llorca and E. Kubala. The Royal Society of Chemistry (London), 1st edn, 2017, **ch. 3**.
53. L. Fielding. NMR methods for the determination of protein-ligand dissociation constants. *Curr. Top. Med. Chem.,* (2003), **3**, 39-53.
54. B. Meyer and T. Peters. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew. Chem. Int. Ed. Engl.,* (2003), **42**, 864-890.
55. M. C. Fernández-Alonso, M. A. Berbís, A. Canales, A. Ardá, F. J. Cañada and J. Jiménez-Barbero. New Applications of High-Resolution NMR in Drug Discovery

and Development. In *New Applications of NMR in Drug Discovery and Development*, eds. L. Garrido and N. Beckmann. The Royal Society of Chemistry, 2013, **ch. 1**.

56. M. P. Williamson, T. F. Havel and K. Wüthrich. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.,* (1985), **182**, 295-315.
57. R. Konrat, I. Burghardt and G. Bodenhausen. Coherence transfer in nuclear magnetic resonance by selective homonuclear Hartmann-Hahn correlation spectroscopy. *J. Am. Chem. Soc.,* (1991), **113**, 9135-9140.
58. B. R. Seavey, E. A. Farr, W. M. Westler and J. L. Markley. A relational database for sequence-specific protein NMR data. *J. Biomol. NMR,* (1991), **1**, 217-236.
59. K. Wüthrich. *NMR of proteins and nucleic acids*. John Wiley & Sons (New York), 1st edn, 1986, pp 130-161.
60. K. Wüthrich. *NMR of proteins and nucleic acids*. John Wiley & Sons (New York), 1st edn, 1986, pp 40-44.
61. M. P. Williamson. Using chemical shift perturbation to characterize ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.,* (2013), **73**, 1-16.
62. F. H. T. Allain, Y. M. Yen, J. E. Masse, P. Schultze, T. Dieckmann, R. C. Johnson and J. Feigon. Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding. *EMBO J.*, (1999), **18**, 2563–2579.
63. M. Sattler, J. Schleucher and C. Griesinger. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients *Prog. Nucl. Magn. Reson. Spectrosc.,* (1999), **34**, 93-158.
64. L. E. Kay, G. Y. Xu and T. Yamazaki. Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation. *J. Magn. Reson. A,* (1994), **109**, 129-133.
65. S. Grzesiek and A. Bax. Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J. Biomol. NMR,* (1993), **3**, 185-204.
66. M. Ikura, L. E. Kay and A. Bax. A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry,* (1990), **29**, 4659-4667.
67. S. Grzesiek and A. Bax. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.,* (1992), **114**, 6291–6293.
68. S. Grzesiek and A. Bax. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.,* (1992), **99**, 201-207.
69. A. Meissner and O. W. Sørensen. Sequential HNCACB and CBCANH protein NMR pulse sequences. *J. Magn. Reson.,* (2001), **151**, 328-331.

70. http://www.bmrb.wisc.edu/ref_info/ The option “Interactive Plots for Protein Amino Acid Chemical Shift Distribution Histograms” gives access to an interactive window where the chemical shift distribution for each amino acid type is shown, compiling the information from all deposited protein structures.
71. D. S. Wishart, B. D. Sykes and F. M. Richards. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.,* (1991), **222**, 311-333.
72. H. Le and E. Oldfield. Correlation between 15N NMR chemical shifts in proteins and secondary structure. *J. Biomol. NMR,* (1994), **4**, 341-348.
73. A. Bundi and K. Wüthrich. ^{1}H-NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers,* (1979), **18**, 285-297.
74. D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges and B. D. Sykes. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR,* (1995), **5,** 67-81.
75. Y. Shen, F. Delaglio, G. Cornilescu and A. Bax. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR,* (2009), **44**, 213-223.
76. M.V. Berjanskii, S. Neal and D.S. Wishart, “PREDITOR: a web server for predicting protein torsion angle restraints”. *Nucleic Acids Res.,* (2006), **34**, W63-69.
77. M. V. Berjanskii and D. S. Wishart. Application of the random coil index to studying protein flexibility. *J. Biomol. NMR,* (2008), **40**, 31-48.
78. J. Morrison, J. C. Yang, M. Stewart and D. Neuhaus. Solution NMR study of the interaction between NTF2 and nucleoporin FxFG repeats. *J. Mol. Biol.,* (2003), **333**, 587–603.
79. A. J. Lowe, F.M. Pfeffer and P. Thordarson. Determining binding constants from 1H NMR titration data using global and local methods: a case study using [n]polynorbornane-based anion hosts. *Supramol. Chem.,* (2012), **24**, 585-594.
80. J. Krishnamoorthy, V. C. K. Yu and Y.-K. Mok. Auto-FACE: an NMR based binding site mapping program for fast chemical exchange protein-ligand systems. *PLoS ONE,* (2010), **5**, e8943.
81. S. B. Shuker, P. J. Hajduk, R. P. Meadows and S. W. Fesik. Discovering high-affinity ligands for proteins: SAR by NMR. *Science,* (1996), **274**, 1531-1534.
82. J. M. Egner, D. R. Jensen, M. D. Olp, N. W. Kennedy, B. F. Volkman, F. C. Peterson, B. C. Smith and R. B. Hill. Development and validation of 2D difference intensity analysis for chemical library screening by protein-detected NMR spectroscopy. *Chembiochem.,* (2018), **19**, 448-458.
83. K. Pervushin, R. Riek, G. Wider and K. WLthrich. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U S A,* (1997), **94**, 12366-12371.

84. P. J. Hajduk, T. Gerfin, J. M. Boehlen, M. Haberli, D. Marek and S. W. Fesik. High-throughput nuclear magnetic resonance-based screening. *J. Med. Chem.,* (1999), **42**, 2315 – 2317.

85. A. P. Duff, K. L. Wilde, A. Rekas, V. Lake and P. J. Holden. Robust high-yield methodologies for (2)H and (2)H/(15)N/(13)C labeling of proteins for structural investigations using neutron scattering and NMR. *Methods Enzymol.*, (2015), **565**, 3-25.

86. K. H. Gardner and L. E. Kay. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.,* (1998), **27**, 357-406.

87. K. Kazimierczuk, J. Stanek, A. Zawadzka-Kazimierczuk and W. Kozminski. Random sampling in multidimensional NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.,* (2010), **57**, 420-434.

88. M. A. McCoy and D. F. Wyss. Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations. *J. Am. Chem. Soc.,* (2002), **124**, 11758–11763.

89. M. Kjaergaard and F. M. Poulsen. Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J. Biomol. NMR,* (2011), **50**, 157–165.

90. D. Stratmann, R. Boelens and A. M. Bonvin. Quantitative use of chemical shifts for the modeling of protein complexes. *Proteins: Struct. Funct. Bioinf.,* (2011), **79**, 2662–2670.

91. P. J. Hajduk, D. J. Augeri, J. Mack, R. Mendoza, J. Yang, S. F. Betz and S. W. Fesik. NMR-based screening of proteins containing 13C-labeled methyl groups. *J. Am. Chem. Soc.,* (2000), **122**, 7898-7904.

92. N. K. Goto, K. H. Gardner, G. A. Mueller, R. C. Willis and L. E. Kay. A robust and cost-effective method for the production of Val, Leu, Ile (δ1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. *J. Biomol. NMR,* (1999), **13**, 369–374.

93. I. Gelis, A. M. Bonvin, D. Keramisanou, M. Koukaki, G. Gouridis, S. Karamanou, A. Economou and C. G. Kalodimos. Structural basis for signalsequence recognition by the translocase motor SecA as determined by NMR. *Cell,* (2007), **131**, 756–769.

94. S. Wiesner, A. A. Ogunjimi, H.-R. Wang, D. Rotin, F. Sicheri, J. L. Wrana and J. D. Forman-Kay. Autoinhibition of the HECT-Type ubiquitin ligase smurf2 through its C2 domain. *Cell,* (2007), **130**, 651–662.

95. M. A. McCornack, D. M. Boren and P. J. LiWang. Glycosaminoglycan disaccharide alters the dimer dissociation constant of the chemokine MIP-1 beta. *Biochemistry,* (2004), **43**, 10090-10101.

96. C. McInnes, S. Grothe, M. O'Connor-McCourt and B. D. Sykes. NMR study of the differential contributions of residues of transforming growth factor alpha to association with its receptor. *Prot. Eng.,* (2000), **13**, 143–147.

97. M. Reibarkh, T. J. Malia and G. Wagner. NMR distinction of single- and multiple-mode binding of small-molecule protein ligands. *J. Am. Chem. Soc.,* (2006), **128**, 2160-2161.

98. A. D. Bain. Chemical exchange in NMR. *Prog. Nucl. Magn. Reson. Spectrosc.,* (2003), **43**, 63-103.

99. S. Weissbach, F. Flegge and T. Peters. Substrate binding drives active-site closing of human blood group B galactosyltransferase as revealed by Hot-Spot labeling and NMR spectroscopy experiments. *ChemBioChem,* (2018), **19**, 970–978.

100. T. Terada, Y. Ito, M. Shirouzu, M. Tateno, K. Hashimoto, T. Kigawa, T. Ebisuzaki, K. Takio, T. Shibata, S. Yokoyama, B. O. Smith, E. D. Laue and J. A. Cooper. Nuclear magnetic resonance and molecular dynamics studies on the interactions of the Ras-binding domain of Raf-1 with wild-type and mutant Ras proteins. *J. Mol. Biol.,* (1999), **286**, 219–232.

101. M. D. Urbaniak, F. W. Muskett, M. D. Finucane, S. Caddick and D. N. Woolfson. Solution structure of a novel chromoprotein derived from apo-Neocarzinostatin and a synthetic chromophore. *Biochemistry,* (2002), **41**, 11731–11739.

102. C. H. Hardman, R. W. Broadhurst, A. R. C. Raine, K. D. Grasser, J. O. Thomas and E. D. Laue. Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy. *Biochemistry,* (1995), **34**, 16596–16607.

103. B. E. Coggins, X. C. Li, A. L. McClerren, O. Hindsgaul, C. R. H. Raetz and P. Zhou. Structure of the LpxC deacetylase with a bound substrate-analog inhibitor. *Nature Struct. Biol.,* (2003), **10**, 645–651.

104. R. A. Williamson, M. D. Carr, T. A. Frenkiel, J. Feeney and R. B. Freedman. Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation. *Biochemistry,* (1997), **36**, 13882–13889.

105. F. H. Schumann, H. Riepl, T. Maurer, W. Gronwald, K.-P. Neidig and H. R. Kalbitzer. Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *J. Biomol. NMR,* (2007), **39**, 275–289.

106. J. Granot. Determination of dissociation constants of 1:1 complexes from NMR data. Optimization of the experimental setup by statistical analysis of simulated experiments. *J. Magn. Reson. (1969),* (1983), **55**, 216-224.

107. M. Arai, J. C. Ferreon and P. E. Wright. Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *J. Am. Chem. Soc.,* (2012), **134**, 3792–3803.

108. C. J. Craven, B. Whitehead, S. K. A. Jones, E. Thulin, G. M. Blackburn and J. P. Waltho. Complexes formed between calmodulin and the antagonists J-8 and TFP in solution. *Biochemistry,* (1996), **35**, 10287–10299.

109. E. L. Kovrigin. NMR line shapes and multi-state binding equilibria. *J. Biomol. NMR,* (2012), **53**, 257-270.
110. L. M. I. Koharudin, A. R. Viscomi, B. Montanini, M. J. Kershaw, N. J. Talbot, S. Ottonello and A. M. Gronenborn. Structure-function analysis of a CVNH-LysM lectin expressed during plant infection by the rice blast fungus *Magnaportheoryzae*. *Structure,* (2011), **19**, 662–674.
111. J. Feeney, J. G. Batchelor, J. P. Albrand and G. C. K. Roberts. Effects of intermediate exchange processes on the estimation of equilibrium constants by NMR, *J. Magn. Reson.,* (1979), **33**, 519-529.
112. "Method for Detecting Biologically Active Compounds from Compound Libraries": T. Peters and B. Meyer. German Pat. No.19649359, Swiss Pat. No. 690695, US Pat. No. 6214561, GB Pat. No. GB2321104 1996.
113. M. Mayer and B. Meyer. Characterization of ligand binding by Saturation Transfer Difference NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.,* (1999), **38**, 1784-1788.
114. T. Diercks, A. S. Infantino, L. Unione, J. Jiménez-Barbero, S. Oscarson and H. J. Gabius. Fluorinated carbohydrates as lectin ligands: Synthesis of OH/F-substituted N-glycan core trimannoside and epitope mapping by 2D STD-TOCSYreF-NMR spectroscopy. *Chem. Eur. J.,* (2018), **24**, 15761-15765.
115. T. Haselhorst, T. Weimar and T. Peters. Molecular recognition of sialyl Lewis(x) and related saccharides by two lectins. *J. Am. Chem. Soc.,* (2001), **123**, 10705-10714.
116. H. Maaheimo, P. Kosma, L. Brade, H. Brade and T. Peters. Mapping the binding of synthetic disaccharides representing epitopes of chlamydial lipopolysaccharide to antibodies with NMR. *Biochemistry,* (2000), **39**, 12778-12788.
117. M. Vogtherr and T. Peters. Application of NMR based binding assays to identify key hydroxy groups for intermolecular recognition. *J. Am. Chem. Soc.,* (2000), **122**, 6093-6099.
118. J. Klages, M. Coles and H. Kessler. NMR-Based screening: a powerful tool in fragment-based drug discovery. *Mol. BioSyst.,* (2006), **2**, 319-331.
119. J. H. Streiff, N. O. Juranic, S. I. Macura, D. O. Warner, K. A. Jones and W. J. Perkins. Saturation Transfer Difference Nuclear Magnetic Resonance spectroscopy as a method for screening proteins for anesthetic binding. *Mol. Pharmacol.,* (2004), **66**, 929-935.
120. S. Monaco, L. E. Tailford, N. Juge and J. Angulo. Differential epitope mapping by STD NMR spectroscopy to reveal the nature of protein-ligand contacts. *Angew. Chem. Int. Ed. Engl.,* (2017), **56**, 15289-15293.
121. M. G. Szczepina, R. B. Zheng, G. C. Completo, T. L. Lowary and B. M. Pinto. STD-NMR studies suggest that two acceptor substrates for GlfT2, a bifunctional galactofuranosyltransferase required for the biosynthesis of *Mycobacterium tuberculosis* arabinogalactan, compete for the same binding site. *ChemBioChem.,* (2009), **10**, 2052-2059.

122. A. Viegas, J. Manso, F. L. Nobrega and E. J. Cabrita. Saturation-Transfer Difference (STD) NMR: a simple and fast method for ligand screening and characterization of protein binding. *J. Chem. Educ.*, (2011), **88**, 990-994.
123. R. Meinecke and B. Meyer. Determination of the binding specificity of an integral membrane protein by saturation transfer difference NMR: RGD peptide ligands binding to integrin alphaIIbbeta3. *J. Med. Chem.,* (2001), **44**, 3059-3065.
124. A. Benie, R. Moser, E. Bäuml, D. Blaas and T. Peters. Virus-ligand interactions: identification and characterization of ligand binding by NMR spectroscopy. *J. Am. Chem. Soc.,* (2003), **125**, 14-15.
125. A. Viegas, N. F. Bras, N. M. F. S. A. Cerqueira, P. A. Fernandes, J. A. M. Prates, C. M. G. A. Fontes, M. Bruix, M. J. Romao, A. L. Carvalho, M. J. Ramos, A. L. Macedo and E. J. Cabrita. Molecular determinants of ligand specificity in family 11 carbohydrate binding modules—An NMR, X-ray crystallography and computational chemistry approach. *FEBS J.,* (2008), **275**, 2524-2535.
126. A. Viegas, A. L. Macedo and E. J. Cabrita. Ligand-macromolecular interactions in drug discovery: Methods and protocols. *Methods Mol. Biol.,* (2010), **572**, 81-100.
127. A. Mallagaray, J. Lockhauserbäumer, G. Hansman, C. Uetrecht and T. Peters. Attachment of norovirus to histo blood group antigens: A cooperative multistep process. *Angew. Chem. Int. Ed.,* (2015), **54**, 12014-12019.
128. C. Rademacher, N. R. Krishna, M. Palcic, F. Parra and T. Peters. NMR experiments reveal the molecular basis of receptor recognition by a Calicivirus. *J. Am. Chem. Soc.,* (2008), **130**, 3669-3675.
129. B. Fiege, C. Rademacher, J. Cartmell, P. I. Kitov, F. Parra and T. Peters. Molecular details of the recognition of blood group antigens by a human norovirus as determined by STD NMR spectroscopy. *Angew. Chem. Int. Ed.,* (2012), **51**, 928 -932.
130. D. Neuhaus and M. P. Williamson. *The nuclear Overhauser effect in structural and conformational analysis*. WILEY-VCH Verlag GmbH & Co. KGaA (Weinheim), 2nd edn, 2000.
131. J. Yan, A. D. Kline, H. Mo, M. J. Shapiro and E. R. Zartler. The effect of relaxation on the epitope mapping by saturation transfer difference NMR. *J. Magn. Reson.,* (2003), **163**, 270-276.
132. M. Mayer and T. L. James. NMR-based characterization of phenothiazines as a RNA binding scaffold. *J. Am. Chem. Soc.,* (2004), **126**, 4453-4460.
133. M. Mayer and B. Meyer. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.,* (2001), **123**, 6108-6117.
134. G. S. Hansman, S. Shahzad-ul-Hussan, J. S. McLellan, G. Chuang, I. Georgiev, T. Shimoike, K. Katayama, C. A. Bewley and P. D. Kwonga. Structural basis for norovirus inhibition and fucose mimicry by citrate. *J. Virol.,* (2012), **86**, 284-292.

135. C. A. Lepre, J. M. Moore and J. W. Peng. Theory and applications of NMR-based screening in pharmaceutical research. *Chem. Rev.,* (2004), **104**, 3641-3675.

136. K. J. Laidler and J. H. Meiser. *Physical Chemistry*. Benjamin/Cummings Pub. Co. (Menlo Park, California), 1982, p. 430.

137. J. Angulo, P. M. Enríquez-Navas and P. M. Nieto. Ligand-receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates. *Chem. Eur. J.,* (2010), **16**, 7803-7812.

138. a) M. Pickhardt, G. Larbig, I. Khlistunova, A. Coksezen, B. Meyer, E. M. Mandelkow, B. Schmidt and E. Mandelkow. Phenylthiazolyl-hydrazide and its derivatives are potent inhibitors of tau aggregation and toxicity in vitro and in cells. *Biochemistry,* (2007), **46**, 10016-10023. b) A. T. Neffe, M. Bilang, I. Gruneberg and B. Meyer. Rational optimization of the binding affinity of CD4 targeting peptidomimetics with potential anti HIV activity. *J. Med. Chem.,* (2007), **50**, 3482-3488.

139. a) H. Sayahi, O. Zimhony, W. R. Jacobs Jr, A. Shekhtman and J. T. Welch. Pyrazinamide, but not pyrazinoic acid, is a competitive inhibitor of NADPH binding to *Mycobacterium tuberculosis* fatty acid synthase I. *Bioorg. Med. Chem. Lett.,* (2011), **21**, 4804-4807. b) H. Sayahi, K. M. Pugliese, O. Zimhony, W. R. Jacobs Jr, A. Shekhtman A and J. T. Welch. Analogs of the antituberculous agent pyrazinamide are competitive inhibitors of NADPH binding to *M. tuberculosis* fatty acid synthase I. *Chem. Biodivers.,* (2012), **9**, 2582-2596.

140. Y. Cheng and W. H. Prusoff. Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.,* (1973), **22**, 3099-3108.

141. V. Jayalakshmi and N. R. Krishna. Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. *J. Magn. Reson.,* (2002), **155**, 106 -118.

142. V. Jayalakshmi and N. R. Krishna. CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities. *J. Magn. Reson.,* (2004), **168**, 36-45.

143. A. Gimeno, L. M. Santos, M. Alemi, J. Rivas, D. lasi, E. Y. Cotrina, J. Llop, G. Valencia, I. Cardoso, J. Quintana, G. Arsequell and J. Jiménez-Barbero. Insights on the interaction between transthyretin and Aβ in solution. A Saturation Transfer Difference (STD) NMR analysis of the role of Iododiflunisal. *J. Med. Chem.,* (2017), **60**, 5749-5758.

144. A. Gimeno, N. C. Reichardt, F. J. Cañada, L. Perkams, C. Unverzagt, J. Jiménez-Barbero and A. Ardá. NMR and molecular recognition of N-Glycans: remote modifications of the saccharide chain modulate binding features. *ACS Chem. Biol.,* (2017), **12**, 1104-1112.

145. V. Jayalakshmi and N. R. Krishna. Determination of the conformation of trimethoprim in the binding pocket of bovine dihydrofolate reductase from a

STD-NMR intensity-restrained CORCEMA-ST optimization. *J. Am. Chem. Soc.,* (2005), **127**, 14080-14084.

146. V. Jayalakshmi, T. Biet, T. Peters and N. R. Krishna. Refinement of the conformation of UDP-galactose bound to galactosyltransferase using the STD NMR intensity-restrained CORCEMA optimization. *J. Am. Chem. Soc.,* (2004), **126**, 8610-8611.
147. F. Ni. Recent developments in transferred NOE methods. *Prog. Nucl. Magn. Reson. Spectrosc.,* (1994), **26**, 517-606.
148. P. Balaram, A. Bothner-By and J. Dadok. Negative nuclear Overhauser effects as probes of macromolecular structure. *J. Am. Chem. Soc.,* (1972), **94**, 4015–4017.
149. M.P. Williamson. The transferred NOE. In *Modern Magnetic Resonance*, ed. G. A. Webb. Springer (Dordrecht), 1st edn, 2008.
150. B. J. Stockman and C. Dalvit. NMR screening techniques in drug discovery and drug design. *Prog. Nucl. Magn. Reson. Spectrosc.,* (2002), **41**, 187-231.
151. G. C. K. Roberts. NMR spectroscopy in structure-based drug design. *Curr. Opin. Biotechnol.,* (1999), **10**, 42-47.
152. B. Meyer, T. Weimar and T. Peters. Screening mixtures for biological activity by NMR. *Eur. J. Biochem.,* (1997), **246**, 705-709.
153. C. B. Post. Exchange-transferred NOE spectroscopy and bound ligand structure determination. *Curr. Opin. Struct. Biol.,* (2003), **13**, 581-588.
154. A. Phan-Chan-Du, M.-C. Petit, G. Guichard, J.-P. Briand, S. Muller, M. T. Cung. Structure of antibody-bound peptides and retroinverso analogues. A transferred nuclear Overhauser effect spectroscopy and molecular dynamics approach. *Biochemistry,* (2001), **40**, 5720-5727.
155. K. Wakamatsu, A. Takeda, T. Tachi and K. Matsuzaki. Dimer structure of magainin 2 bound to phospholipid vesicles. *Biopolymers,* (2002), **64**, 314-327.
156. H. Furukawa, T. Hamada, M. K. Hayashi, T. Haga, Y. Muto, H. Hirota, S. Yokoyama, K. Nagasawa and M. Ishiguro. Conformation of Ligands bound to the muscarinic acetylcholine receptor. *Mol. Pharmacol.,* (2002), **62**, 778-787.
157. P. L. Jackson, H. N. B. Moseley and N. R. Krishna. Relative effects of protein-mediated and ligand-mediated spin-diffusion pathways on transferred NOESY, and implications on the accuracy of the bound ligand conformation. *J. Magn. Reson. Ser. B,* (1995), **107**, 289-292.
158. A. P. R. Zabell and C. B. Post. Intermolecular relaxation has little effect on intra-peptide exchange-transferred NOE intensities. *J. Biomol. NMR,* (2002), **22,** 303-315.
159. J. Zheng and C. B. Post. Protein indirect relaxation effects in exchange-transferred NOESY by a rate-matrix analysis. *J. Magn. Reson.,* (1993), **101**, 262-270.
160. E. Z. Eisenmesser, A. P. R. Zabell and C. B. Post. Accuracy of bound peptide structures determined by exchange transferred nuclear Overhauser data: a simulation study. *J. Biomol. NMR,* (2000), **17**, 17–32.

161. S. P. Arepalli, C. P. Glaudemans, G. D. Daves, P. Kovac and A. Bax. Identification of protein-mediated indirect NOE effects in a disaccharide-Fab' complex by transferred ROESY. *J. Magn. Reson. Ser. B,* (1995), **106**, 195-198.
162. J. L. Asensio, F. J. Cañada and J. Jiménez-Barbero. Studies of the bound conformations of methyl alpha-lactoside and methyl beta-allolactoside to ricin B chain using transferred NOE experiments in the laboratory and rotating frames, assisted by molecular mechanics and dynamics calculation. *Eur. J. Biochem.,* (1995), **233**, 618-630.
163. N. Murali, G. K. Jarori, S. B. Landy and B. D. N. Rao: Two-dimensional transferred Nuclear Overhauser effect spectroscopy (TrNOESY) studies of nucleotide conformations in creatine kinase complexes: effects due to weak nonspecific binding. *Biochemistry,* (1993), **32**, 12941-12948.
164. J. B. Jordan, L. Poppe, X. Xia, A. C. Cheng, Y. Sun, K. Michelsen, H. Eastwood, P. D. Schnier, T. Nixey and W. Zhong. Fragment based drug discovery: Practical implementation based on 19F NMR spectroscopy. *J. Med. Chem.,* (2012), **55**, 678-687.
165. A. Vulpetti, U. Hommel, G. Landrum, R. Lewis and C. Dalvit. Design and NMR-based screening of LEF, a library of chemical fragments with different local environment of fluorine. *J. Am. Chem. Soc.,* (2009), **131**, 12949–12959.
166. C. Dalvit. Ligand- and substrate-based ^{19}F NMR screening: Principles and applications to drug discovery. *Prog. Nucl. Magn. Reson. Spectrosc.,* (2007), **51**, 243–271.
167. C. Dalvit. Theoretical analysis of the competition ligand-based NMR experiments and selected applications to fragment screening and binding constant measurements. *Concepts Magn. Reson.,* (2008), **32A**, 341-372.
168. S. Kiihne and R. G. Bryant. Protein-bound water molecule counting by resolution of (1)H spin-lattice relaxation mechanisms. *Biophys. J.,* (2000), **78**, 2163–2169.
169. J. W. Peng, J. Moore and N. Abdul-Manan. NMR experiments for lead generation in drug discovery. *Prog. Nucl. Magn. Reson. Spectrosc.,* (2004), **44**, 225-256.
170. C. Dalvit and A. Vulpetti. Ligand-based fluorine NMR screening: principles and applications in drug discovery projects. *J. Med. Chem.,* (2019), **62**, 2218-2244.
171. C. Dalvit, E. Ardini, G. P. Fogliatto, N. Mongelli and M. Veronesi. Reliable high-throughput functional screening with 3-FABS. *Drug Discov. Today,* (2004), **9**, 595-602.
172. C. Dalvit, D. T. Hadden, R. W. Sarver, A. M. Ho and B. J. Stockman. Multi-selective one dimensional proton NMR experiments for rapid screening and binding affinity measurements. *Comb. Chem. High Throughput Screen.,* (2003), **6**, 445-453.
173. J. M. Moore. NMR techniques for characterization of ligand binding: utility for lead generation and optimization in drug discovery. *Biopolymers,* (1999), **51**, 221-243.

174. P. J. Hajduk, E. T. Olejniczak and S. W. Fesik. One-dimensional relaxation- and Diffusion-edited NMR methods for screening compounds that bind to macromolecules. *J. Am. Chem. Soc.,* (1997), **119**, 12257-12261.
175. B. Boulat, R. Konrat, I. Burghardt and G. Bodenhausen. Measurement of relaxation rates in crowded NMR spectra by selective coherence transfer. *J. Am. Chem. Soc.,* (1992), **114**, 5412–5414.
176. M. J. Moure, A. Eletsky, Q. Gao, L. C. Morris, J. Y. Yang, D. Chapla, Y. Zhao, C. Zong, I. J. Amster, K. W. Moremen, G. J. Boons and J. H. Prestegard. Paramagnetic tag for glycosylation sites in glycoproteins: Structural constraints on heparan sulfate binding to Robo1. *ACS Chem. Biol.,* (2018), **13**, 2560−2567.
177. W. Jahnke, S. Ruedisser and M. Zurini. Spin label enhanced NMR screening. *J. Am. Chem. Soc.,* (2001), **123**, 3149-3150.
178. W. Jahnke, L. B. Perez, C. G. Paris, A. Strauss, G. Fendrich and C. M. Nalin. Second-Site NMR screening with a spin-labeled first ligand. *J. Am. Chem. Soc.,* (2000), **122**, 7394-7395.
179. S. L. Cobb and C. D. Murphy. ^{19}F NMR applications in chemical biology. *J. Fluor. Chem.,* (2009), **130**, 132-143.
180. M. Veronesi, F. Giacomina, E. Romeo, B. Castellani, G. Ottonello, C. Lambruschini, G. Garau, R. Scarpelli, T. Bandiera, D. Piomelli and C. Dalvit. Fluorine nuclear magnetic resonance-based assay in living mammalian cells. *Anal. Biochem.,* (2016), **495**, 52-59.
181. J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, M. Rance and N. J. Skelton. *Protein NMR Spectroscopy: Principles and Practice*. Elsevier Academic Press (Boston), 2nd edn, 2007.
182. C. Dalvit, P. E. Fagerness, D. T. Hadden, R. W. Sarver and B. J. Stockman. Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability. *J. Am. Chem. Soc.,* (2003), **125**, 7696-7703.
183. H. M. McConnell. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.,* (1958), **28**, 430-431.
184. T. J. Swift and R. E. Connick. NMR-relaxation mechanisms of O^{17} in aqueus solutions of paramagnetic cations and the lifetime of water molecules in the first coordination sphere. *J. Chem. Phys.,* (1962), **37**, 307–320.
185. C. Dalvit, N. Mongelli, G. Papeo, P. Giordano, M. Veronesi, D. Moskau and R. Kümmerle. Sensitivity improvement in 19F NMR-based screening experiments: theoretical considerations and experimental applications. *J. Am. Chem. Soc.,* (2005), **127**, 13380-13385.
186. T. C. Farrar and E. D. Becker. *Pulse and Fourier Transform NMR, Introduction to Theory and Methods.* Elsevier Academic Press (New York), 1st edn, 1971.
187. E. L. Hahn. Spin Echoes. *Phys. Rev.,* (1950), **80**, 580–594.
188. H. Y. Carr and E. M. Purcell. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.,* (1954), **94**, 630-638.

189. S. Meiboom and D. Gill. Modified Spin-Echo method for measuring nuclear relaxation times. *Rev. Sci. Instrum.,* (1958), **29**, 688-691.
190. Z. Luz and S. Meiboom. Nuclear Magnetic Resonance study of the protolysis of trimethylammonium ion in aqueous solution—Order of the reaction with respect to solvent. *J. Chem. Phys.,* (1963), **39**, 366-370.
191. J. D. Martínez, P. Valverde, S. Delgado, C. Romanò, B. Linclau, N. C. Reichardt, S. Oscarson, A. Ardá, J. Jiménez-Barbero and F. J. Cañada. Unraveling Sugar Binding Modes to DC-SIGN by Employing Fluorinated Carbohydrates. *Molecules,* (2019), **24**, E2337.
192. W. Jahnke, M. J. J. Blommers, C. Fernández, C. Zwingelstein and R. Amstutz. Strategies for the NMR-based identification and optimization of allosteric protein kinase inhibitors. *ChemBioChem,* (2005), **6**, 1607-1610.
193. J. D. Taylor, P. J. Gilbert, M. A. Williams, W. R. Pitt and J. E. Ladbury. Identification of novel fragment compounds targeted against the pY pocket of v-Src SH2 by computational and NMR screening and thermodynamic evaluation. *Proteins,* (2007), **67**, 981-990.
194. A. J. Doerr, M. A. Case, I. Pelczer and G. L. McLendon. Design of a functional protein for molecular recognition: specificity of ligand binding in a metal-assembled protein cavity probed by ^{19}F NMR. *J. Am. Chem. Soc.,* (2004), **126**, 4192-4198.
195. W. S. Price. Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion: Part 1. Basic theory. *Concepts Magn. Reson.,* (1997), **9**, 299-336.
196. W. S. Price, P. W. Kuchel and B. A. Cornell. Microviscosity of human erythrocytes studied with hypophosphite and ^{31}P-NMR. *Biophys. Chem.,* (1989), **33**, 205-215.
197. W. S. Price and L.-P. Hwang. Some recent developments in NMR approaches for studying liquid molecular dynamics and their biological applications. *J. Chin. Chem. Soc. (Taipei),* (1992), **39**, 479-496.
198. W. S. Price. Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion: Part II. Experimental aspects. *Concepts Magn. Reson.,* (1998), **10**, 197-237.
199. V. Gilard, S. Trefi, S. Balayssac, M.-A. Delsuc, T. Gostan, M. Malet-Martino, R. Martino, Y. Prigent and F. Taulelle. DOSY NMR for drug analysis. In *NMR spectroscopy in pharmaceutical analysis,* eds. H. Ulrike, W. Iwona and D. Bernd. Elsevier (Amsterdam), 1st edn, 2008, **ch. 6**.
200. Y. Cohen, L. Avram and L. Frish. Diffusion NMR spectroscopy in supramolecular and combinatorial chemistry: an old parameter--new insights. *Angew. Chem., Int. Ed.,* (2005), **117**, 520-554.
201. E. O. Stejskal and J. E. Tanner. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *J. Chem. Phys.,* (1965), **42**, 288-292.

202. J. E. Tanner. Use of the stimulated echo in NMR diffusion studies. *J. Chem. Phys.,* (1970), **52**, 2523-2526.
203. D. Burstein. Stimulated echoes: Description, applications, practical hints. *Concepts Magn. Reson.,* (1996), **8**, 269-278.
204. D. Wu, A. Chen and C. S. Johnson Jr. An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses. *J. Magn. Reson. Ser. A,* (1995), **115**, 260-264.
205. A. Jerschow and N. Müller. Suppression of convection artifacts in stimulated-echo diffusion experiments. Double-stimulated-echo experiments. *J. Magn. Reson.,* (1997), **125**, 372-375.
206. N. Esturau, F. Sánchez-Ferrando, J. A. Gavin, C. Roumestand, M.-A. Delsuc and T. Parella. The use of sample rotation for minimizing convection effects in self-diffusion NMR measurements. *J. Magn. Reson.,* (2001), **153**, 48-55.
207. R. Huo, R. Wehrens, J. van Duynhoven and L. M. C. Buydens. Assessment of techniques for DOSY NMR data processing. *Anal. Chim. Acta,* (2003), **490**, 231-251.
208. C. S. Johnson Jr. Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Prog. Nucl. Magn. Reson. Spectrosc.,* (1999), **34**, 203-256.
209. T. Brand, E. J. Cabrita and S. Berger. Intermolecular interaction as investigated by NOE and diffusion studies. *Prog. Nucl. Magn. Reson. Spectrosc.,* (2005), **46**, 159-196.
210. B. Antalek. Using pulsed gradient spin echo NMR for chemical mixture analysis: How to obtain optimum results. *Concepts Magn. Reson.,* (2002), **14**, 225-258.
211. J. M. Newman and A. Jerschow. Improvements in complex mixture analysis by NMR: DQF-COSY iDOSY. *Anal. Chem.,* (2007), **79**, 2957-2960.
212. A. Jerschow and N. Muller. 3D Diffusion-ordered TOCSY for slowly diffusing molecules. *J. Magn. Reson. Ser. A,* (1996), **123**, 222-225.
213. A. S. McLachlan, J. J. Richards, A. R. Bilia and G. A. Morris. Constant time gradient HSQC-iDOSY: practical aspects. *Magn. Reson. Chem.,* (2009), **47**, 1081-1085.
214. A. Gierer and K. Z. Wirtz. *Naturforsch. A.,* (1953), **8**, 522-532.
215. H.-C. Chen and H.-S. Chen. Diffusion of crown ethers in alcohols. *J. Phys. Chem.,* (1984), **88**, 5118-5121.
216. A. M. Krause-Heuer, N. J. Wheate and J. Aldrich-Wright. Diffusion-based studies on the self-stacking and nanorod formation of platinum (II) intercalators. *Chem. Commun.,* (2009), 1210-1212.
217. C. Monteiro and C. Hervé du Penhoat. Translational diffusion of dilute aqueous solutions of sugars as probed by NMR and hydrodynamic theory. *J. Phys. Chem. A,* (2001), **105**, 9827-9833.
218. E. J. Cabrita and S. Berger. High-resolution DOSY NMR with spins in different chemical surroundings: influence of particle exchange. *Magn. Reson. Chem.,* (2001), **66**, S142-148.

219. P. Groves, M.O. Rasmussen, M.D. Molero, E. Samain, F. J. Cañada, H. Driguez and J. Jiménez-Barbero. Diffusion ordered spectroscopy as a complement to size exclusion chromatography in oligosaccharide analysis. *Glycobiology,* (2004), **14**, 451-456.

220. P. Groves, M. Palczewska, M. D. Molero, G. Batta. F. J. Cañada and J. Jiménez-Barbero. Protein molecular weight standards can compensate systematic errors in diffusion-ordered spectroscopy. *Anal. Biochem.,* (2004), **331**, 395-397.

221. A. Chen and M. J. Shapiro. Affinity NMR. *Anal. Chem.,* (1999), **71**, 669A-675A.

222. C. S. Johnson Jr. Effects of chemical exchange in diffusion-ordered 2D NMR spectra. *J. Magn. Reson. Ser. A,* (1993), **102**, 214-218.

223. E. J. Cabrita, S. Berger, P. Brauer and J. Karger. High-resolution DOSY NMR with spins in different chemical surroundings: Influence of particle exchange. *J. Magn. Reson.,* (2002), **157**, 124-131.

CHAPTER 4: DC-SIGN RECOGNITION OF THE HISTO BLOOD GROUP ANTIGENS A AND B

Carbohydrates are undoubtedly the most abundant biomolecules in nature. In living organisms, these compounds play diverse roles that remain essential for the maintenance of homeostasis. Namely, they serve as energy source (glucose metabolism), as structural components of the cell wall, or as markers for ensuring the proper folding of newly synthesized polypeptides and the stability, solubility and trafficking of the final glycoproteins [1-2]. Additionally, as exemplified in *Chapter 1*, glycans are essential at the surface of all types of cells due to their role of molecular recognition elements in the establishment of cell-cell or host/pathogen dialogs, as well as many other processes including cell proliferation and differentiation [3-4].

From a structural point of view, simple carbohydrates work as building blocks in the assembly of highly complex glycan structures. Similar to proteins and nucleic acids, the number of "letters" that form this "sugar code" is limited to less than 15 monosaccharides [5-6]. However, the particular structural features of these molecules, such as the existence of two possible anomers and the functionalization with multiple hydroxyl groups (4 available in the most common pyranoses, Glc, Gal and Man), enormously increases the versatility of such compounds as essential bricks in nature [7]. Such a degree of complexity is actually observed in living cells, although not all the possible combinations are really found due to the limited availability and specificity of the glycosyltransferases in charge of catalyzing the formation of new glycosidic bonds [8]. The additional complexity comes then from other aspects regarding the accessibility of the glycan to the receptor and its conformational flexibility, which both finely determine the correct glycan presentation that makes the sugar-receptor interaction feasible [6]. Also indirectly, these properties influence the recognition event by affecting the enthalpy-entropy compensation already known for sugars [9]. Overall, glycan flexibility governs multiple factors concerning the affinity and significance of a given epitope, for instance, it determines how high the entropic penalty will be due to desolvation or loss of conformational freedom [10-11], or whether a terminal sialic acid moiety is an adequate substrate for a particular receptor depending on to which position of the preceding sugar is attached (e.g. the α2-3- or the more flexible α2-6 linkage) [12]. An additional level of complexity is finally reached when natural glycans are jointly considered as taking part in the extensive network of varied sugars that coats the cell membrane. This "sugary coating of cells" is commonly named as "glycocalix" [13-14], and all the existing glycoproteins and sugar-containing entities are perfectly organized and their density is strictly regulated [15]. Thus, they work in consonance with

their corresponding receptors to maintain a fine equilibrium that allows the cell to suitably display unique molecular patterns, thereby communicating with the neighboring cells as well as with cells from the innate and adaptive immune system [16]. In fact, both an altered sugar composition and an abnormal amount of certain sugar receptors are known hallmarks of cancer and autoimmune diseases [17-18], and hence a potential source of biomarkers and targets for biotherapeutics [19].

As mentioned, glycosyltransferases (GnT) [20] are the enzymes responsible for elongating and branching the initial heptasaccharide precursor $Man_5GlcNAc_2$ in mammals [21-22], increasing their glycan diversity. Usually, these GnTs carry out the elongations by attaching successive LacNAc disaccharides (Galβ1-4GlcNAc), and eventually, the branch is capped with a terminal characteristic motif [7,23]. Since the most accessible positions of these long sugar chains are commonly the terminal caps, these are often the target of specific protein receptors, and hence can be considered as "self" signatures. Among these "self" motifs, the most prominent and studied ones are sialic acid and histo blood group antigens. Sialic acid is a single monosaccharide recognized by diverse receptors, including host proteins like siglecs [24] and pathogen receptors like hemagglutinin from influenza virus [25]. In contrast, the term "histo blood group antigen" is applied to a series of several antigens that share a common feature: terminal fucosylation.

Usually, two distinctions are done, namely the ABO antigens and the Lewis-related antigens, although all of them are biologically generated from the same initial precursor (Figure 4.1A). Such a precursor is a disaccharide unit composed of Gal and Glu that is intercepted by specific fucosyl- (FT) and sialyltransferases (ST) [26]. The final architecture of a given histo blood group antigen will then depend on the specific set of GnTs that the individual possesses, which will determine the initial precursor (at least six types described), and the subsequent modifications driven on it. Thus, due to GnT polymorphisms [27], glycan composition and self-motifs vary among different individuals of the same species, what can be understood as a useful natural self-protection strategy to guarantee the surveillance of a certain population against carbohydrate-mediated infections or pathologies.

As defining structural features, ABO antigens are α1-2-fucosylated, whereas Lewis-type antigens can be either α1-3 or α1-4-fucosylated (Figure 4.1A). Moreover, after the first fucosylation, the terminal Gal may be fucosylated

again (Lewis B/Y types) or alternatively sialylated (sialyl-Lewis A/X types) [26]. In contrast, ABO antigen biosynthesis is always ended up after the introduction of another Gal residue on the core Gal, mediated by specific galactosyltransferases [28-29].

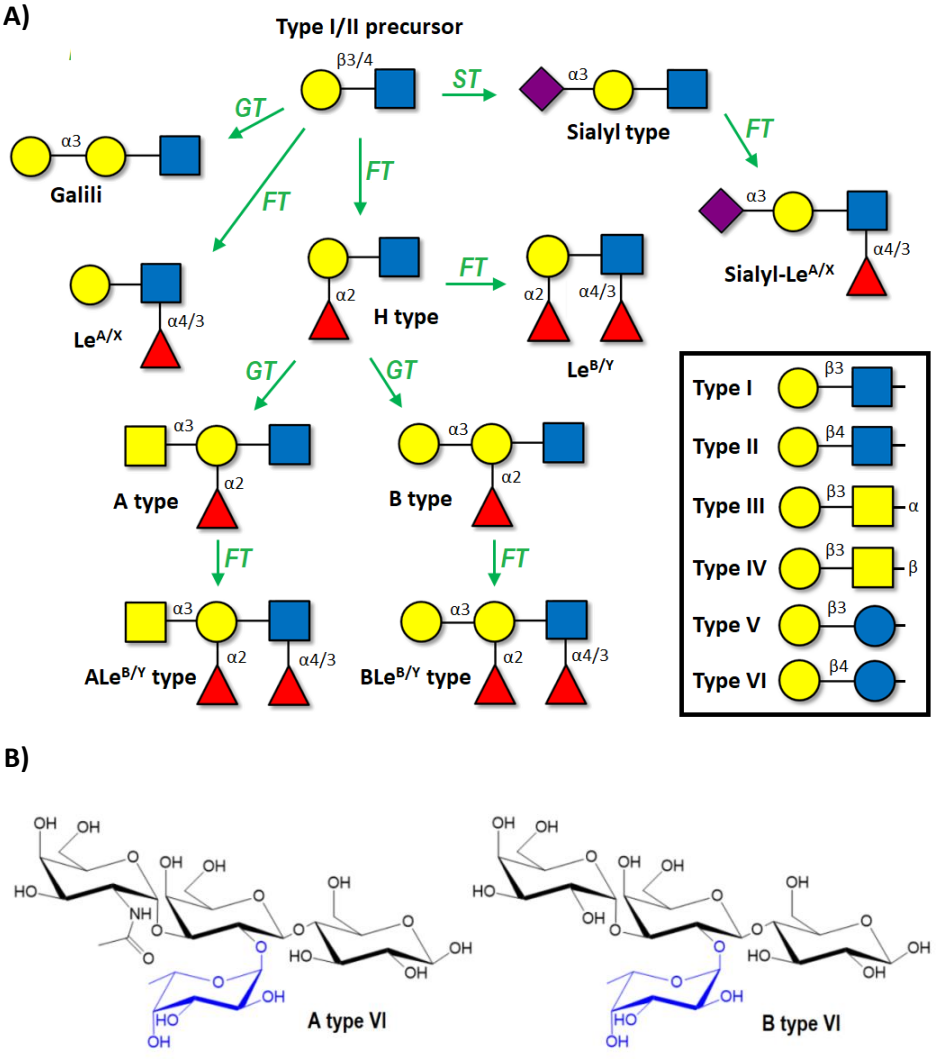


Figure 4.1. *A) Biosynthetic routes that give rise to the different histo blood group antigens found in mammalian cells. For each particular transformation, the corresponding GnT is indicated in green. FT: fucosyltransferase; GT: galactosyltransferase; ST: sialyltransferase. On the right, the six types of disaccharide precursors are represented. B) Structures of the blood group antigens A and B type VI used in this work.*

The following chapter compiles most of the results achieved along this dissertation, focused on studying the recognition of the histo blood group

antigens A and B (Figure 4.1B). Particularly, the tetrasaccharides used herein have their core Gal linked to a reducing Glc through a β1-4 bond, whereby they are both type VI antigens. The common denomination "blood group" applied to these structures was firstly coined by Karl Landsteiner, an Austrian physician and immunologist who firstly realized that the blood agglutination resulting from mixing blood samples provided by two different patients was provoked by the presence of specific molecular patterns, which he identified and named as blood group antigens (three different, A, B and O) [30]. Soon, these antigens were discovered to be also in other tissues like gut, skin, respiratory tract and even neurons; and in biological fluids such as saliva [26]. In this regard, the prevalence of the classical name "blood group antigens" simply comes from the circumstances concerning their discovery and their particular usefulness in transfusion medicine.

4.1. Receptor production

Nuclear magnetic resonance allows the study of molecular recognition phenomena in solution. Hence, for the studies carried out herein, a soluble form of the DC-SIGN protein had to be firstly afforded in a suitable buffered medium, supplemented with the corresponding oligosaccharides of interest. Given that DC-SIGN is a membrane protein, the full protein sequence only exists as a correctly folded and functional structure in the presence of a bilayer membrane constituted by amphipathic lipids. According to the existing literature [31-33], carbohydrate-protein interactions involving this kind of lectins are always investigated by exclusively making use of the extracellular soluble portions. Specifically, the carbohydrate recognition domain (CRD) constitutes the minimum relevant fragment needed to undertake molecular recognition studies, since in most of cases this region has its own structural integrity and its folding does not depend on the neck regions. In addition, the expression of the full extracellular domain (ECD: CRD + neck) is a common strategy usually employed as well. The main advantage of this latter method is that the oligomeric state of the protein can be maintained, as the neck regions often take part in the corresponding oligomerization processes. Thus, the native state of the receptor of interest is better reproduced as well as it also enables the study of other events affecting the binding, as clustering effects, for instance. Herein, both the CRD and the ECD of DC-SIGN were produced by protein expression protocols in living systems, in enough amounts for performing the NMR experiments. Additionally, the CRD was also expressed in isotopically enriched media to prepare samples uniformly labeled either with

only ^{15}N or with ^{15}N and ^{13}C, which served for performing the corresponding heteronuclear NMR experiments.

4.1.1. Protein expression

The DNA fragments containing the CRD and ECD sequences were respectively inserted in a pET15b plasmid (Thermo Fischer Scientific). The CRD fragment comprised 151 amino acids, from His254 to Ala404, while the ECD sequence included 185 additional residues which constituted the neck segment, starting from Glu70 [34]. To note, in the native protein N80 is glycosylated, although the biological role of such a glycosylation is still unclear [35]. In either case, the presence of glycans at that position does not influence the structural stability of the ECD, and therefore there was not any drawback associated to the expression of this construct in bacterial media, completely devoid of glycosylation machinery. Apart, the CRD construct was initially designed with an additional His-Tag fragment linked to the N-termini (MGSSHHHHHHSSGLVPRGSHML). In principle, the purpose of this modification was to test whether a faster and more straightforward purification of the CRD was possible by means of a simple Ni^{2+}-containing His-Trap. Unfortunately, the CRD could not be efficiently purified and other options were considered, as subsequently detailed.

The corresponding plasmids were amplified in *E. Coli* cells (DH5α strain) by thoroughly following the standard protocol specified in the QIAprep® Spin Miniprep Kit (250), using the suitable buffers included in the kit. The final concentrations of the isolated plasmids were variable but always ranging around 100-150 ng/µL, as confirmed by UV absorbance measured at 260 nm. With these amounts, 2-3 µL of the plasmid were enough to inoculate a small aliquot containing BL21/DE3 cells. After the thermal shock at 42º, the transformed cells were incubated in LB medium at 37ºC to obtain competent colonies. Then, precultures of 100 mL in LB medium were prepared from single colonies, and grown overnight at 37ºC in the presence of 100 mg/l of ampicillin. Those precultures were then diluted up to 1 L of final culture in LB at 37ºC, and supplemented with 100 mg/L ampicillin. Under continuous shaking (180 rpm), the cells were allowed to grow until reaching an optical density of 0.60-0.65 (600 nm). At that moment, the protein expression was induced by adding 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) and the exponential growth was prolonged for 2.5 to 3 hours, maintaining a constant temperature of 37ºC. Following this methodology, the final ODs achieved were about 4.0-5.0 for each liter prepared. These results were similar for the CRD and for the ECD. Eventually, the produced cells were

harvested at lower temperature (4ºC) by centrifugation, and the resulting pellets were resuspended in the minimal amount of Tris 10 mM buffer (pH = 8) and stored at -80ºC.

The preparation of the isotopically labeled CRD was carried out through a different protocol. After obtaining competent BL21 cells expressing DC-SIGN, small pre-inocules of 5-10 mL were prepared in LB from single colonies. These precultures were grown overnight at 37ºC under gentle shaking. Afterwards, each preculture was directly diluted into 1 liter of sterile M9 minimal medium containing 100 mg/L ampicillin, as well as other supplements (biotin, thiamine) and necessary ions (Ca^{2+}, Mg^{2+}, Na^{+}, Cl^{-}, SO_4^{2-}…). As expected, the cell growth was significantly slower for two reasons: the low starting OD (10 mL of competent cells diluted into 1 L) and the intrinsically harder cell development in a M9 medium, containing limited nutrients. Even so, ODs close to 0.40-0.45 could be reached within 5-7 hours at 37ºC. This methodology permitted to skip the intermediate overnight preculture in 100 mL of M9, such that the uniformly labeled protein could be efficiently expressed in two days as much. Of course, the initial amount of LB pre-inocule (without labeled components) had to be very small in comparison with the final volume, thereby guaranteeing a percentage of isotopic labeling above 98-99%. In contrast to the normal methodology, this protocol was proved to be rather inefficient when the induced cultures were grown for few hours at 37º C. Thus, after reaching ODs between 0.40-0.45, the temperature was decreased to 24 ºC, and the cell induction with IPTG was carried out as soon as the OD surpassed 0.60, maintaining the constant shaking at every moment. Then, the exponential growth was allowed to progressing overnight at 24 ºC, thereby maximizing the protein overexpression and minimizing the adverse effects of protease-mediated degradations. By means of this protocol, the final ODs achieved were slightly lower (about 3.0-3.5), but still large enough for the preparation of NMR samples.

4.1.2. Protein purification

The purification protocol used herein was similar to those already reported in previous works [36-37], with some little modifications to adapt the protein production to the available equipment and to the final purpose of the prepared samples. For both the CRD and the ECD, the applied methodology was equivalent, and it basically consisted of two steps of protein separation and isolation through chromatography techniques, preceded by a refolding process using dialysis.

First, the stored pellets were thawed and quickly resuspended in a larger volume of Tris 10 mM, always maintaining the temperature below 4ºC and the pH at 7.8-8.0. To extract the produced protein from the bacteria, the cells were immediately lysed by sonication. Complete cell lysis was efficiently achieved after 18 cycles of sonication (30 seconds each, t_{burst} = 1 sec., Amp. = 60%) leaving 1 minute of resting between consecutive cycles. Then, the soluble components of the cell debris were subsequently separated from the insoluble fraction by ultracentrifugation at 30000 rpm (1 hour, 4º C). By SDS-PAGE (4-12%), the protein was confirmed to remain unfolded at the insoluble fraction of the sonicated pellets, forming inclusion bodies as previously reported [36,38]. Once the inclusion bodies were isolated, the construct was primarily solubilized as completely unfolded linear peptide chains, in the presence of strong denaturing conditions, namely, urea 6 M and β-mercaptoethanol 10 mM (~0.05% v/v). To note, the pH value along all the purification steps was always set to 8.0, using Tris.HCl to buffer the solution. The pellets were manually smashed and ground in a glass mortar while applying the denaturing buffer. Specifically, at least 8 mL of the solution were added per liter of cell culture present in the pellet. Once finished the resuspension (almost homogeneous opaque solution), the resulting mixture was quickly sonicated again (4-5 additional cycles using the same setup) and gently stirred overnight at low temperature, at least for 15-20 hours. After several tests, these guidelines were checked out to be optimal for recovering the maximal amount of protein from the inclusion bodies (Figure 4.2, *cols. 1-2*). In particular, the sonication steps (also after the solubilization in urea) were key to extract notable amounts of protein from the cellular media and largely increased the final yields, whereby the duration of this stage was always thoroughly controlled by SDS-PAGE evaluation of the retired fractions. Once the solubilization was achieved, the residual insoluble cell components were sedimented by ultracentrifugation at 40000 rpm for 2.5-3.0 hours (4 ºC), and the supernatant layer containing the soluble protein was carefully decanted. In this regard, those liquid fractions located on the bottom of the centrifuge tube were preventively discarded, since they showed an increased viscosity as a result of high concentrations of nucleic acids, as checked by UV absorbance (260 nm). The subsequent dialysis in the presence of these impurities was noticed to somehow hamper the progressive folding of the protein, instead favoring intense re-precipitation phenomena and giving rise to an important loss of soluble protein. In these cases, when the residual pellet still displayed a marked amount of protein, the solubilization was repeated again to improve the protein yields.

The final supernatants contained the soluble unfolded protein mixed with other soluble biomolecules from *E. Coli*. The liquid fractions collected were all placed into a semipermeable membrane, and stepwise dialyzed in three different buffers (pH = 8.0) containing decreasing amounts of urea (4 M, 2 M and 0 M respectively). Each dialysis step was performed for 24 hours at 4 ºC, under gentle stirring and against 5 L of dialysis buffer. Importantly, the structural integrity and functionality of the protein required the presence of fixed amounts of Ca^{2+} (10 mM) and β-mercaptoethanol (0.01% v/v) throughout the entire refolding process [39]. A moderate ionic strength (150 mM NaCl) was also mandatory to hamper undesired aggregation phenomena promoted by the hydrophobic parts of the protein, which will constitute the core of the folded scaffold [40]. Under these conditions, the protein could be successfully folded into its functional three-dimensional shape after three days, and neither chaperones nor disulfide isomerases were further needed along this purification step [41]. To highlight, other authors have efficiently accomplished the protein folding in the presence of glutathione as molecular system to adjust the redox potential [38]. However, in our particular conditions, the β-mercaptoethanol provided the impetus to form all the disulfide bonds present in the CRD (three) and in the ECD (four) constructs.

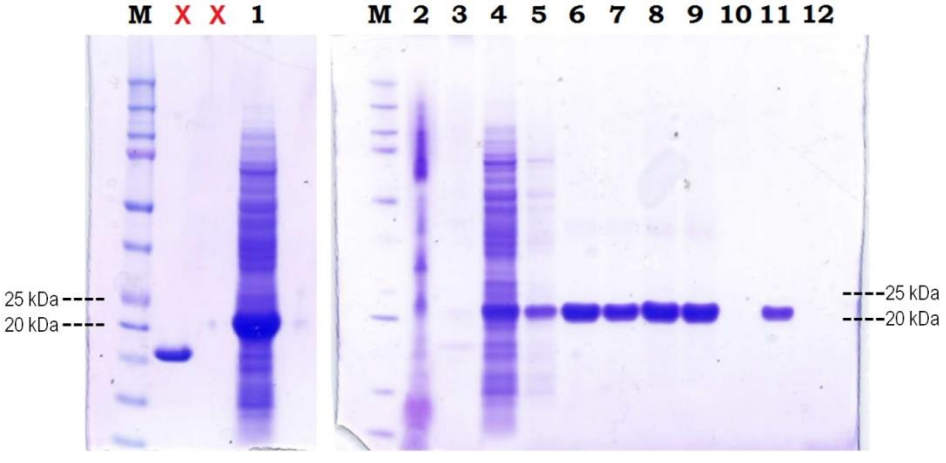


Figure 4.2. *Purification steps monitored by 4-12% SDS-PAGE (1% β-mercaptoethanol) for the ^{15}N-labeled CRD (from 12 L of M9 culture). Column legend: M) Marker. 1) Supernatant containing the solubilized inclusion bodies (6 M urea). 2) Remaining insoluble pellet in 6 M urea. 3) Residual pellet precipitated from the dialyzed solution of the protein. 4) Flow-through collected during the protein loading into the Man-sepharose column. 5) Column wash (>100 mL) after protein loading. 6) to 9) First four fractions eluted from the affinity column (5 mL EDTA 10 mM). 10) Fraction collected in the 15th elution (>75 mL EDTA). 11) Fractions collected from the AKTA containing the pure monomer. 12) Flow-through during the concentration steps (negative control).*

During the last dialysis step, a residual precipitate was always observed, which presumably corresponded to the remaining aggregated unfolded proteins and other smaller components, whose solubility depended on the urea. Thus, whereas the times for the intermediate dialysis steps were rather flexible (around 20-30 hours), the last step (0 M urea) had to be carefully controlled, since the prolonged exposition of the functional protein to the unfolded components of the precipitate favored the progressive loss of the already folded protein. In this scenario, the dialysis membrane was routinely removed after a reasonable time (16-24 hours) while the percentage of insoluble precipitate was low enough yet (Figure 4.2, *col. 3*). In all cases, this precipitate was immediately separated by ultracentrifugation at 25000 rpm (30 min, 4 ºC) and discarded.

The folded functional protein was subsequently purified by affinity chromatography, taking advantage of the natural ability of DC-SIGN to selectively recognize Man. Following the methodology already described [36-37], a mannose-sepharose column was employed to separate DC-SIGN from the rest of soluble *E. Coli* biomolecules. The solution containing the folded protein was mixed with 10 mL of mannose-sepharose beads previously coated with the loading buffer (25 mM Tris-HCl, 150 mM NaCl, 10 mM $CaCl_2$, pH = 8), and the resulting mixture was vigorously stirred overnight to ensure the maximal percentage of protein attachment. Afterwards, the stationary phase was allowed to depositing at the bottom of the cartridge, and the upper solution layer was fully passed through without applying vacuum or pressure. The beads were then washed with the same loading buffer (at least 100 mL) to remove any other protein attached in a non-specific fashion. In many cases, the amount of protein incubated with the beads was higher than the limit concentration that these beads could harbor. Then, the collected flow-throughs and washes were stored at 4º C for subsequent loadings after checking, by SDS-PAGE, that they still contained important amounts of non-retained protein (Figure 4.2, *cols. 4-5*). Once loaded and separated from the starting mixture, the DC-SIGN construct was eluted from the stationary phase by disrupting the Ca^{2+}-dependent recognition of Man with EDTA 10 mM. The elution was performed stepwise by adding small amounts of the elution buffer, up to 5 mL (25 mM Tris-HCl, 150 mM NaCl, 10 mM EDTA, pH = 8). After each addition, the column was gently shaken for 15-20 min, and then the buffer was passed through and collected in fractions of about 1 mL. The protein concentration in these fractions was individually checked by UV absorbance at 280 nm. The first fractions often provided lower yields, since the added EDTA firstly sequestered the free calcium ions present in the

loading buffer that coated the column beads. From this point, the amount of extracted protein exponentially decreased along successive elutions, and the process was stopped when the measured absorbances were close to the detection limit of the device (A = 0.05-0.10) (Figure 4.3). Then, the elution was considered as finished. The analyses of the recorded absorbances corroborated that the amount of extracted protein surely ranged within 96-98%. After several purifications, these columns were proved to harbor a total amount of protein that could be eluted with ca. 50-60 mL of EDTA 10 mM (10-12 elutions of 5 mL) (Figures 4.3 and 4.2, *cols. 6-9*). Additionally, the efficiency was checked to slightly increase when the added volumes of chelating agent were smaller than 5 mL (same number of elutions but a lower EDTA total volume).

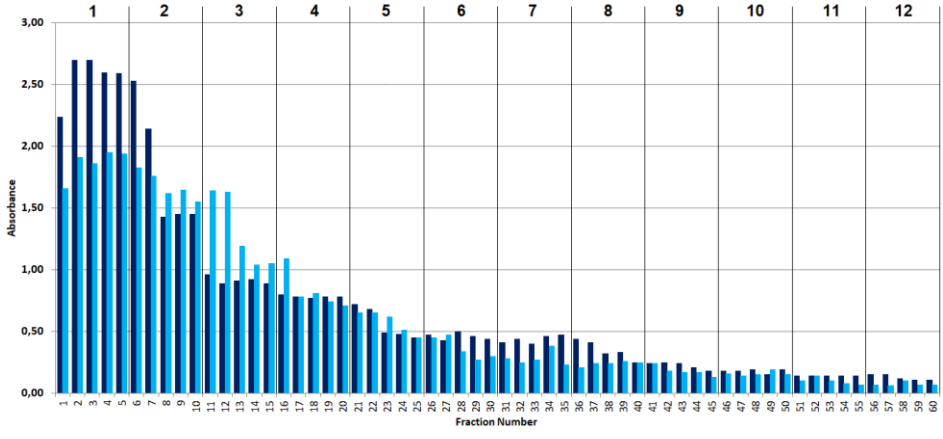


Figure 4.3. *Absorbances measured at 280 nm (ε_λ = 52855 M^{-1} cm^{-1}, estimated from ProtParam) for the 60 first fractions collected along 12 successive elutions with EDTA 10 mM (5 mL each).The displayed data correspond to the eluted fractions from the first protein loading (dark bars) and the subsequent re-loading of the remaining protein contained in the flow-throughs (light bars).*

Finally, the purity and the size of the isolated protein were assessed by size exclusion chromatography (SEC). Due to the partial misfolding experienced by the protein in the absence of Ca^{2+}, a small amount of the protein was detected to form fine white threads, as a result of total unfolding. Although the percentage of lost protein was minimal (undetectable changes in the UV absorbance) these colloidal particles were preventively removed by centrifugation at 4000 rpm (4 ºC), thereby avoiding a synergic and faster progression of the unfolding process. For the SEC run, dextran-based columns (Superdex) were used for both constructs, CRD and ECD. Importantly, as previously reported [41], the protein cannot be eluted in native conditions

when these columns are employed in the presence of Ca^{2+}. The stationary phase is basically composed of dextran, a polymeric carbohydrate containing Gal and Glc. Even though DC-SIGN exhibits a weak affinity for these sugars, the interaction is strong enough to preclude the elution of the protein, and it remains bound to the column in any case (Figure 4.5). Then, the collected fractions from the Man-sheparose column were directly injected to the Superdex column and eluted in the presence of EDTA as well (25 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH = 8). Given the large total volume of the mixed fractions (50-60 mL), the most diluted ones (below A = 1) were primarily concentrated in membrane filters by centrifugation at 2500 rpm (4 ºC). The consensus absorbance for the injected samples was set within the 1-2 range (below 50 µM), since higher protein concentrations in the absence of calcium favored a faster appearance of precipitated protein.

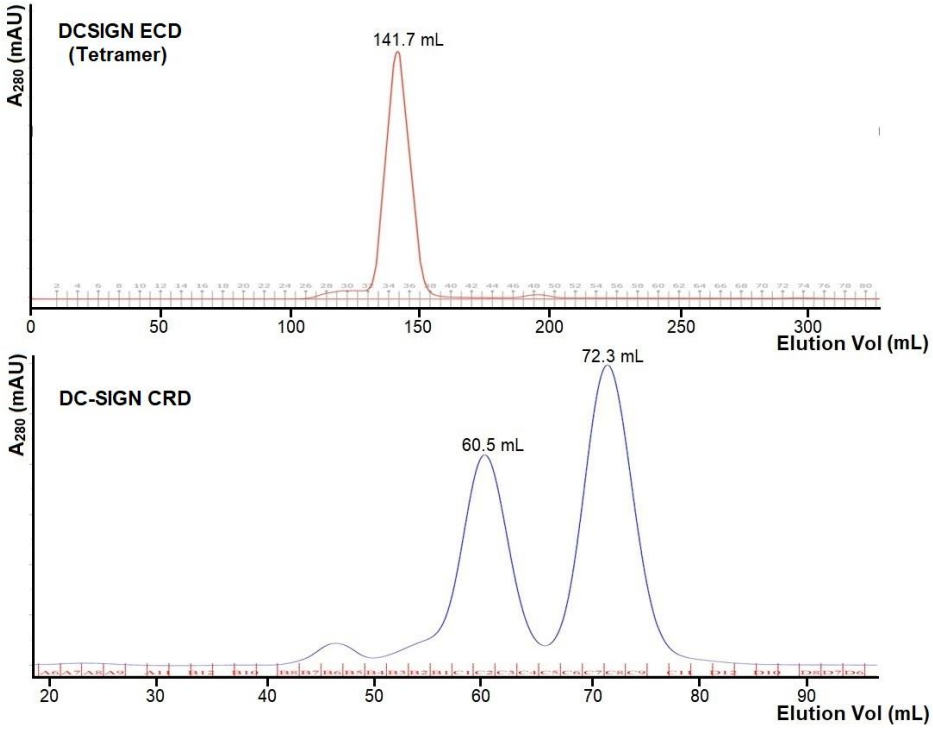


Figure 4.4. *SEC elution profiles for the ECD (Top, Superdex 200, flow rate = 2.5 mL/min) and the CRD (Bottom, Superdex 75, flow rate = 1 mL/min).*

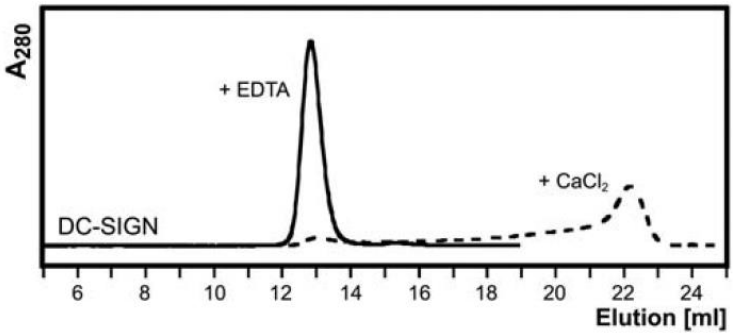


Figure 4.5. *Gel filtration profiles (Tris-HCl, pH = 7.4) obtained for the DC-SIGN CRD when eluted in the presence of either 1 mM EDTA (solid line) or 1 mM $CaCl_2$ (dashed line). Adapted from [41].*

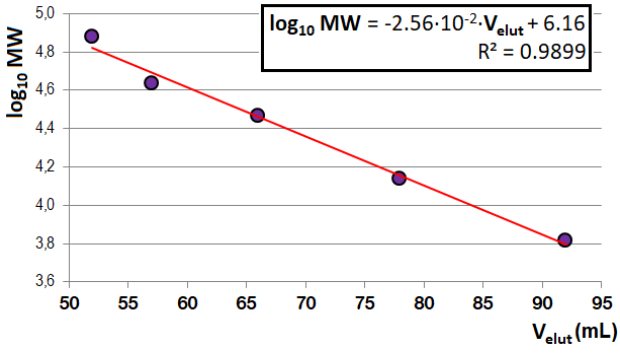
The ECD was eluted in a Superdex 200 column, giving a single elution peak at a retention volume of 141 mL, which corresponded to the self-assembled tetramer (Figure 4.4, *top*). The corresponding fractions containing the pure tetramer were collected and mixed for further concentration. The scenario was remarkably different for the CRD, which was eluted through a Superdex 75 column. In particular, the successive runs constantly allowed the visualization of two main elution peaks, appearing at retention volumes of 60 and 72 mL, respectively (Figure 4.4, *bottom*). Interestingly, the second elution peak, corresponding to the smallest biomolecule, was always the most intense, while the other one was rather variable and usually displayed a concentration between 25-75% of the major component. The molecular sizes associated to these elution volumes were evaluated by means of a linear regression built from five standard patterns already provided by the column manufacturer (Figure 4.6A). The estimated size for the major peak was 20.6 kDa, reasonably close to the expected molecular weight of the DC-SIGN monomer (19.9 kDa). Surprisingly, the other elution peak at 59.9 mL displayed a size of about 41,4 kDa, close to twice the molecular weight of the monomer, what suggested that this second component corresponded to a CRD dimer. At first, it was unknown why the formation of these dimers was possible, since the DC-SIGN CRD has been actually reported to exclusively exist as a monomer in the total absence of any amino acid repeat from the neck [35]. However, taking into account that the designed construct included an odd number of cysteine residues, the possibility that the remaining free cysteine was involved in the formation of disulfide bonds bridging two CRDs was considered. In fact, the counterpart cysteine of C384 was C253, which was not present in the expressed sequence. Both residues are located on the terminal unstructured regions of the CRD, and hence, presumably accessible to the solvent. This

hypothesis was quickly assessed by SDS-PAGE using a marker without β-mercaptoethanol to preserve the cystin groups, which actually enabled the observation of two migration bands, instead of one (Figure 4.6B). Even so, the observation of two elution peaks was significantly reduced but not completely solved in the presence of β-mercaptoethanol (0.01%). In fact, several further tests showed that pure fractions of the monomer tended to reconstitute the dimer again when re-injected in the AKTA system. Certainly, the presence of EDTA somehow favored this process as well, since the ratio dimer/monomer markedly decreased as the time lapse between the elution from the Man-sepharose column and the SEC run was reduced, whereby the time that the CRD was under non-native conditions was crucial. These findings suggested that the partial deconstruction of the loop regions in the absence of Ca^{2+} led to the unspecific association of the monomers, similar to the association phenomena observed for any unfolded protein. Therefore, an equilibrium monomer-dimer was being established and it seemed to be contributed by diverse factors, including the presence of a free cysteine in the construct (C384).

A)

	V_{elut} (mL)	MW (Da)*	log MW
Aprotinin	92	6500	3.813
Ribonuclease A	78	13700	4.137
Carb. Anhydrase	66	29000	4.462
Ovoalbumin	57	43000	4.633
Canalbumin	52	75000	4.875
Peak A	59.9	**41444**	4.617
Peak B	71.7	**20672**	4.315

**Aprox. values provided by the manufacturer.*



B)

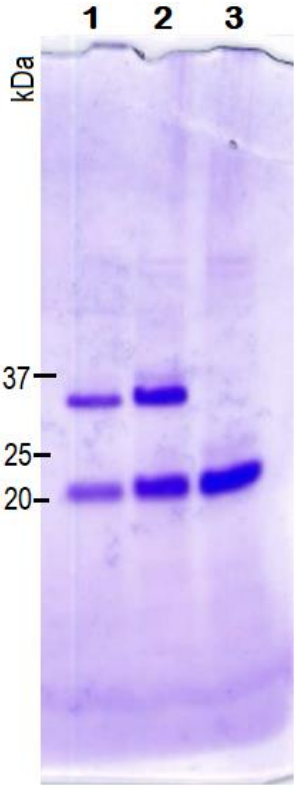


Figure 4.6. *A) Verification of the molecular weights expected for the two species reported along the purification of the CRD construct in the Superdex 75 column. B) Fractions extracted from the Man-sepharose chromatography analyzed by SDS-PAGE in different conditions. Col. 3: fraction loaded with the normal buffer (1% β-mercaptoethanol). Cols. 1&2: Fractions loaded in the absence of reducing agents after thermal denaturation.*

The nature of the described monomer/dimer equilibrium was further analyzed by NMR to try setting a series of experimental guidelines that could minimize the dimerization event during the interaction studies with ligands. Specifically, the ratio monomer/dimer was analyzed through Diffusion Ordered SpectroscopY (2D-DOSY) starting from a sample that remained several days at room temperature, in the sole presence of Ca^{2+} 4 mM, without either reducing agents or EDTA. Under such conditions, a 2D-DOSY spectrum of the sample was acquired at 310 K and its diffusion coefficient was determined from the indirect dimension. In parallel, the corresponding diffusion coefficients of three commercial proteins (BSA, Carbonic Anhydrase and Lysozyme) were estimated in the same conditions (310 K, buffer Tris with Ca^{2+} 4mM and pH = 8) through independent 2D-DOSY experiments. The measured D coefficients, together with their respective known molecular weights, were used to build a regression function that correlated the experimental diffusion coefficients with the corresponding molecular sizes [42-43]. Then, the average molecular weight of the previous sample of DC-SIGN was interpolated through this mathematical function, and the resulting value suggested that the ratio monomer/dimer was about 40:60 (Figure 4.7).

Interestingly, when the 2D-DOSY was recorded again in the presence of 10 mM Man, the sample size was noticed to diminish below 24 kDa, what was indicative of an 80-85% of monomer in solution. Eventually, the addition of a little excess (0.5 mM) of dithiothreitol (DTT) sharply shifted the diffusion coefficient to higher values, suggesting the almost exclusive presence of the monomer and only marginal percentages of the dimer. On the basis of these findings, and given that the oxidation/reduction process was the major event contributing to this dimerization phenomenon, all further NMR measurements were always performed in the presence of at least 2 mM DTT to avoid the existence of several species with different sizes and minimize the impact of the monomer/dimer equilibrium in the protein degradation. Although less pronounced, it is also worth noting that the carbohydrate binding event seemed to play a role in the self-association process, presumably improving the rigidity of the long loop close to the binding site and precluding the unspecific association that could be driven by these flexible regions.

	MW (Da)	Log_{10} MW	Log_{10} D (2D-DOSY)	Fraction of monomer
Lysozyme	14300	4.1553	-9.76	-
Carb. Anhydrase	29000	4.4624	-9.88	-
Bovine Serum Albumin	66200	4.8208	-10.06	-

DC-SIGN	**32522**	4.5122	-9.91	**0.37**
DC-SIGN + 10 mM Man	**23252**	4.3665	-9.85	**0.84**
DC-SIGN + 0.5mM DTT	**18485**	4.2668	-9.80	**~1.0**

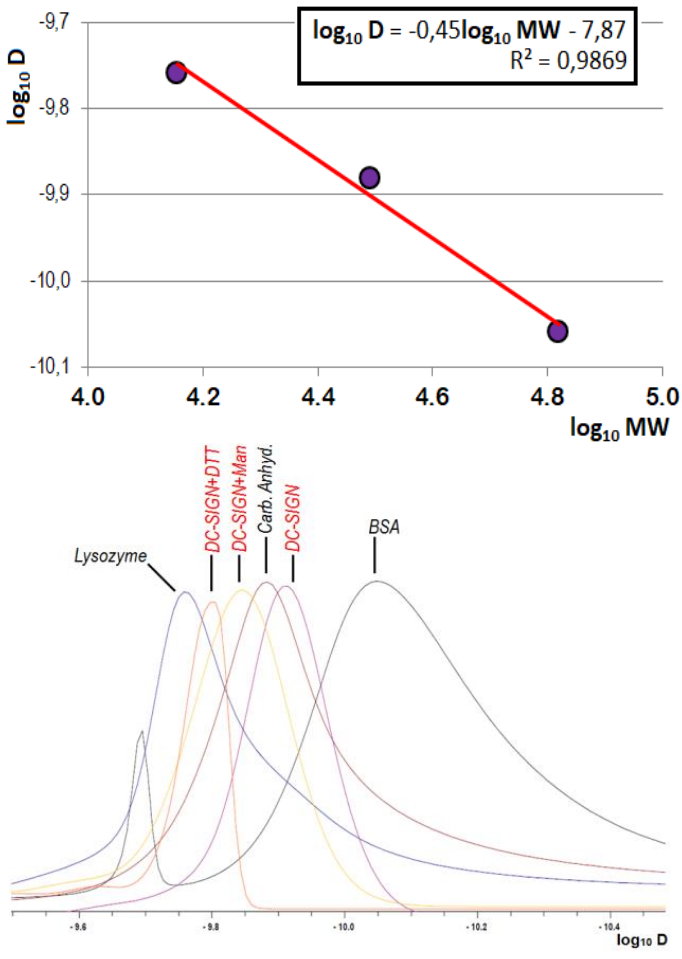


Figure 4.7. *2D-DOSY-based calculation of the average molecular weight of the CRD construct under different conditions. The experimental values of the diffusion coefficients (D) were taken from the vertical projection of the 2D-DOSY (D-based chromatogram), as shown on the bottom figure. The ratios monomer/dimer (last column) are estimated through the real molecular weights related to the monomer (19990 Da) and the dimer (39980 Da).*

The pure protein fractions collected from the AKTA system were mixed and concentrated by centrifugation at 2000-2500 rpm (4 ºC) in a membrane filter. Then, the protein was washed within the same filter with a non-deuterated buffer containing Tris-HCl and NaCl (pH = 8), as well as 4 mM $CaCl_2$ and 2 mM DTT, thereby restoring the Ca^{2+} and removing the EDTA. At least four washes were performed in which a dilution factor of 1:10 was applied each time, thus leaving a residual concentration of EDTA below 1 µM in all cases. Eventually, the buffer was changed to the suitable NMR conditions following the same guidelines, and making use of deuterated compounds (25 mM Tris-d_{11}, 150 mM NaCl, 4 mM $CaCl_2$, 2 mM DTT-d_{10}, pH = 8). For the CRD assignment and the ^{15}N-HSQC-based titrations with the uniformly labeled samples, the buffer was prepared in H_2O/D_2O 9:1, enabling the observation of the exchangeable NH groups, while the experiments from the ligand point of view were better recorded in the presence of high purity D_2O (99.99%, $[H]_{min}$ = 10 mM). As described above, the buffer change took four concentration stages again, thereby applying a total dilution factor of 10^4 to the initial non-deuterated components, which reduced their concentrations below 5 µM and efficiently erased their NMR signals from the proton spectrum. The final pure and concentrated protein was eventually distributed in aliquots of a given concentration according with the planned experiments, and stored at -80 ºC.

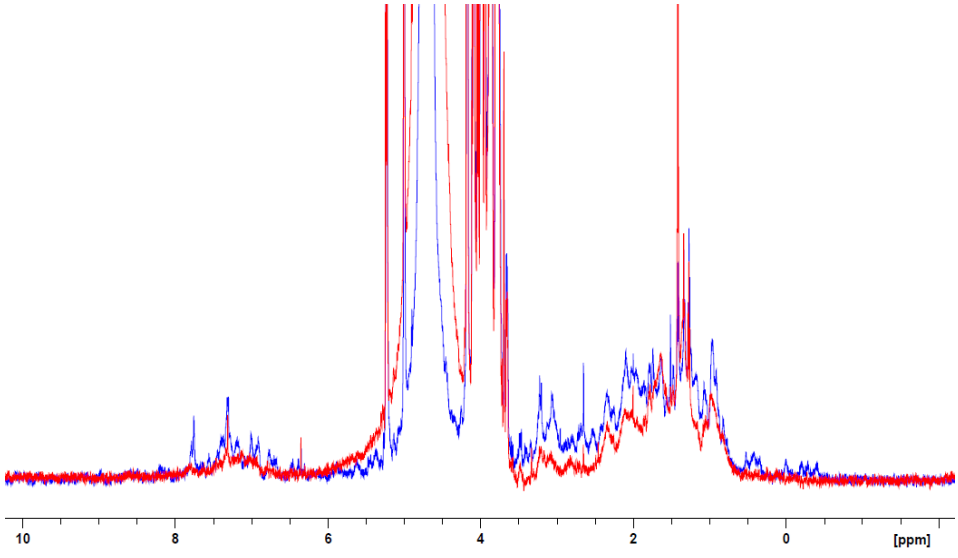


Figure 4.8. *Control ^{1}H-1D spectra of the purified monomeric CRD recorded at 310 K (25 mM Tris-d_{11}, 150 mM NaCl, 4 mM $CaCl_2$, 2 mM DTT-d_{10}, pH = 8) in the presence of 50 equivalents of Man_3 (blue), and after adding 10 mM EDTA to the same sample (red).*

In average, the global overexpression and purification processes afforded 1.5-1.7 mg of protein per liter of grown culture. Routinely, the integrity and folding of the protein were verified by NMR after finishing the purification, checking the peak broadening of the protein profile in the ^{1}H-NMR 1D spectrum (except for those peaks from Tris or DTT, if visible) and specially monitoring the presence of several nuclei shifted below 0 ppm, which typically correspond to the aliphatic sidechains (Ile, Val, Leu) located on the inner parts of a folded scaffold (Figure 4.8). These peaks quickly disappeared after adding EDTA to the sample, suggesting severe changes in the protein folding when Ca^{2+} is absent. Also, the binding abilities of the purified DC-SIGN constructs were tested by STD NMR. Two STDs were acquired at 310 K in the presence of a large excess of the Man_3 trisaccharide, a known ligand of this lectin. Either under aliphatic or aromatic saturation, an STD profile could be observed when the lectin was into the NMR tube, thus confirming the carbohydrate binding ability of the used construct (Figure 4.9).

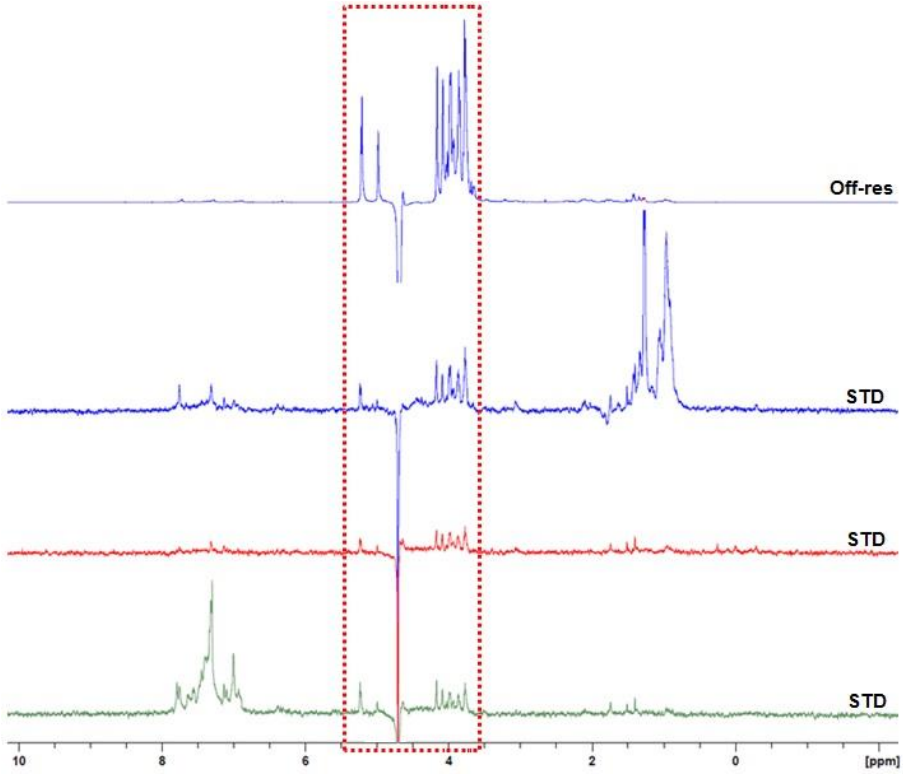


Figure 4.9. *STD profiles obtained for the Man_3 trisaccharide at 310 K in the presence of the DC-SIGN CRD (t_{sat} = 2 sec.). As displayed, the ligand could be saturated (red dashed box) by applying the selective pulse at three different resonances (0.96 ppm (blue), 0.00 ppm (red), 7.31 ppm (green)), thereby confirming the interaction of the oligosaccharide with the expressed construct.*

4.2. CRD backbone assignment

To study the interaction of the blood group antigens with DC-SIGN from the receptor point of view, the CRD rather than the extracellular domain (ECD) was used to make the relaxation mechanisms T_1 and T_2 less efficient and then maximize the signal-to-noise ratio along the data acquisition [44]. Using the ^{15}N-labeled CRD samples, a series of quick tests were performed to choose the proper work temperature. At first, the studies performed by the groups of Prestegard [38] and Dixon [32], respectively using DC-SIGN and L-SIGN, were chosen as references. In both cases, the assignment and subsequent titrations were conducted at physiological temperature: 37º C. This is, in principle, the correct choice for getting realistic results in binding analyses, but not mandatory for the CRD assignment. In the latter case, it is better giving priority to seek for optimal acquisition conditions during the three-dimensional NMR experiments.

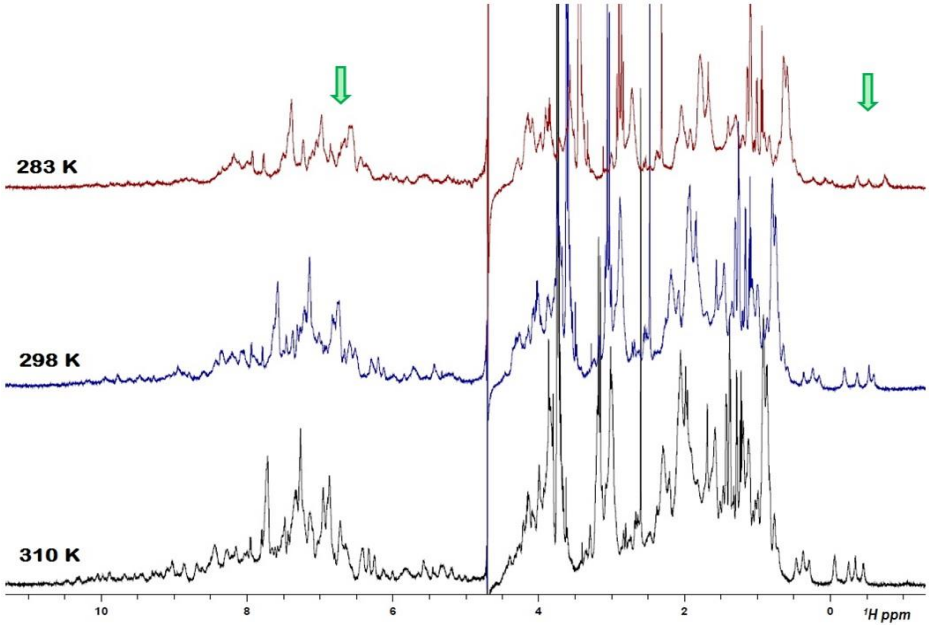


Figure 4.10. *Overview of the ^{1}H-1D spectra of DC-SIGN CRD recorded at three different temperatures. The regions where the signal broadening is easily seen are indicated by green arrows.*

Both ^{1}H-1D and ^{15}N-HSQC experiments showed a progressive signal broadening as the temperature was decreased from 310 K (37ºC). In the 1D spectra (Figure 4.10), this was specially noticed for the aliphatic resonances around 0 ppm and the NH region between 6-8 ppm. The ^{15}N-HSQC at 298 K (25ºC) displayed multiple weakened peaks (Figure 4.11), and surprisingly, at

283 K (10ºC) most of the CRD backbone peaks were no longer visible, except for those from the disordered protein regions in the middle of the spectrum. Thus, local protein motions in folded regions were noticed to get slower, moving from a fast-exchange regime to an intermediate-regime in which many signals were notably reduced due to coalescence phenomena. Consequently, the three-dimensional CRD assignment was conducted at 310 K.

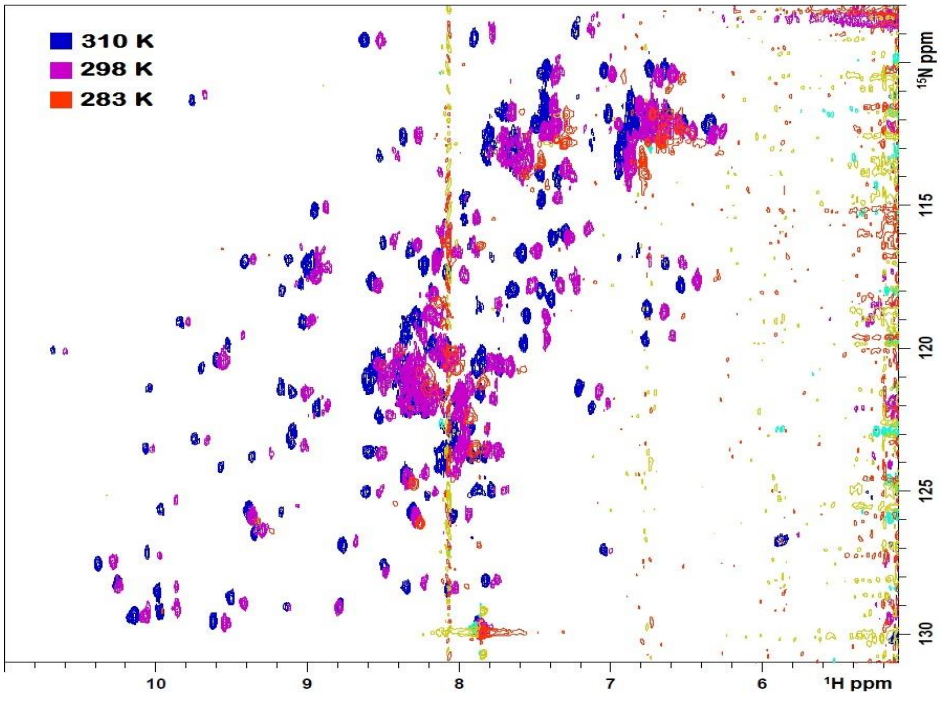


Figure 4.11. *Superimposition of the ^{15}N-HSQC spectra of DC-SIGN CRD acquired at 310 K, 298 K and 283 K, respectively.*

The group of Prestegard had already deposited the backbone assignment of the DC-SIGN CRD in the BMRB database, with the accession code 19931 [38,45]. However, at the first sight, striking differences were detected between our protein fingerprint and some peak positions reported in the deposited file (Figure 4.12). Namely, several expected correlation peaks did not appear in our experiment (yellow diamonds), and conversely, we visualized additional spin systems not reported before (red dashed circles). The only differences between Prestegard´s conditions and ours were the concentrations of $CaCl_2$ (2,5 mM/4 mM) and NaCl (100 mM/150 mM). In this scenario, a higher ionic strength might be triggering slight conformational changes, thus explaining such discrepancies. It should be mentioned that many of these cases involved residues from the N- and C-termini, highly

flexible regions whose disordered structure is rather unpredictable, so attention was mainly paid to explain the most relevant cases.

The CRD backbone assignment was achieved by employing the typical set of 3D NMR spectra to gather information on ^{1}H, ^{15}N, $^{13}C\alpha$ and $^{13}C\beta$ nuclei resonances. These included HN(CO)CA, HNCA, HN(CO)CACB and HNCACB. In particular, the intensity of the correlation peaks in the HNCA experiment come essentially from the carbon adjacent to the nitrogen (HN_i-$C\alpha_i$), whereas the magnetization is poorly transferred to the $^{13}C\alpha_{i-1}$ [46-47]. Therefore, the ^{15}N-HSQC projection of this experiment was firstly used to quickly detect the Trp, Asn and Gln sidechains, since their peak intensities are expected to dramatically drop due to the lack of an adjacent Cα (Figure 4.13). The Trp sidechains usually appear at lower fields than the rest of NH crosspeaks (orange box on the bottom left corner) and the opposite happens for Asn/Gln sidechains (purple boxes on the top right corner). All the Trp NH groups were identified (nine), whereas the identification of the Asn/Gln terminal amides was limited by the intense signal overlapping. To note, a TROSY pulse sequence was used to record the ^{15}N-HSQC spectra and thus reduce signal crowding by increasing the resolution.

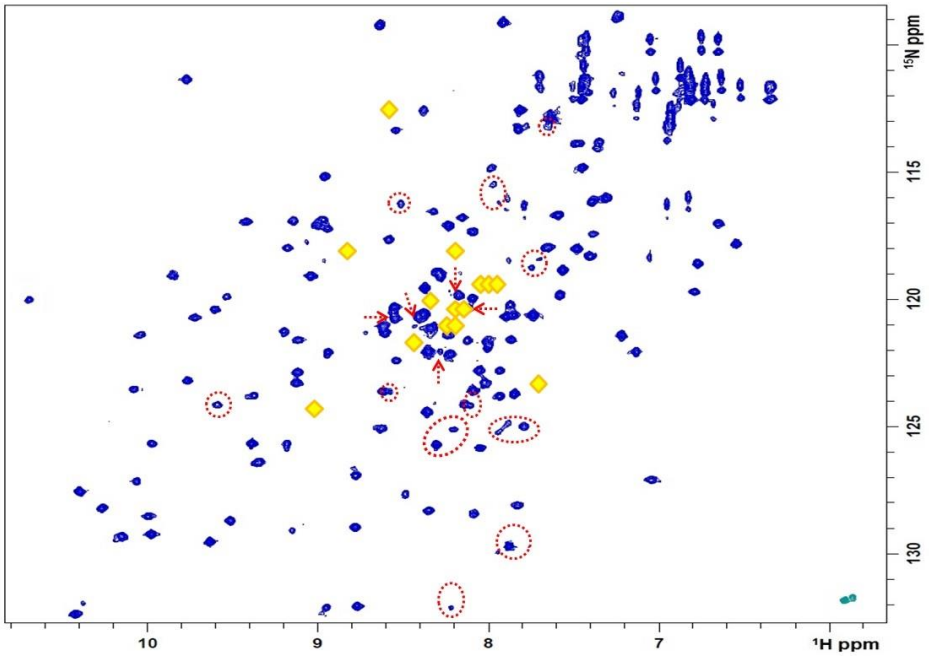


Figure 4.12. *Main discrepancies found between the existing deposited assignment and our experimental ^{15}N-HSQC (TROSY). Yellow diamonds: expected peaks (BMRB) that do not appear in our experiments; red circles and arrows: systems unlinked to published data.*

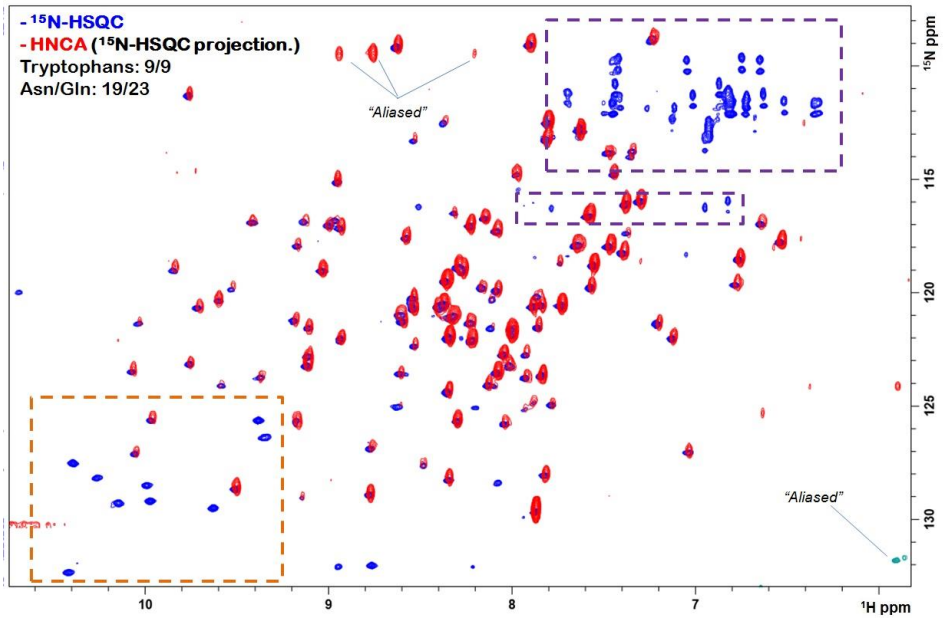


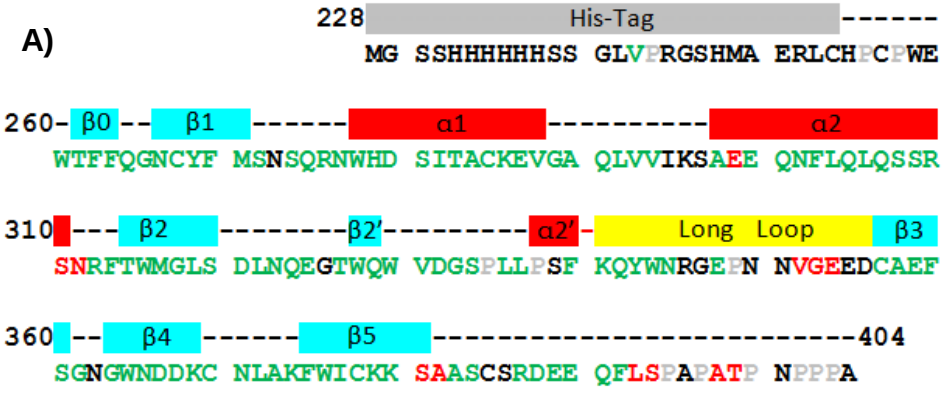
Figure 4.13. *Identification of the HN-containing sidechains of Trp, Asn and Gln residues through superimposition of the routine ^{15}N-HSQC (TROSY) with the same projection extracted from the 3D-HNCA. As noted, sidechains only show their crosspeaks on the original 2D (blue), but not on the 3D-derived plane (red). Also, signal splitting can be noticed for Asn/Gln sidechain correlations in the nitrogen dimension, as a result of the different evolution of the NH and NH_2 correlations according to the TROSY sequence [48].*

Overall, the 3D assignment of the free CRD was unsuccessful at first, as a great deal of lost signals were noticed, especially those from Cβ nuclei. Indeed, the relaxation mechanisms were observed to be so efficient that nearly all Cβ correlation peaks were missing during the acquisition of the HNCACB experiment. As a result, a first assignment was attempted by simultaneously using the information from ^{13}Cα-^{13}Cα-1 pairs and the deposited assignments to establish coherent sequential correlations. At last, the gathered information was roughly the same (Table on Figure 4.14) although the poor signal dispersion of the Cα resonances precluded the sequential connection of all found spin systems, whereby 14 systems remained with ambiguous or unclear assignments (Figure 4.14).

As previously described in *Section 4.1*, the CRD stability was noticed to increase in two scenarios: after adding a reducing agent (DTT) and after binding to a natural ligand. Although not fully understood, it was clear that ligand binding improves the stability of the long loop, somewhat favoring the presence of smaller species (CRD monomers) that showed sharper NMR

signals. Thus, the entire CRD assignment was repeated in the presence of a large excess of a natural ligand, being D-Man the chosen candidate. The experiments were recorded with a double ^{13}C-^{15}N-labeled sample of the CRD at 225 uM, in the presence of 30 mM of D-Man (133 equivalents) to ensure protein saturation.

A)



B)

	BMRB 19931	Free CRD	Man-bound CRD
%Assign	118/140 **(84%)**	104/140 **(74%)**	120/140 **(86%)**
Cα	139/140 (99%)	119/140 (85%)	124/140 (88%)
Cα-1	-	119/140 (85%)	126/140 (90%)
Cβ	119/131 (91%)	2/131 (1.5%)	110/131 (84%)
Cβ-1	-	50/131 (36%)	111/131 (85%)

Total spin systems (residues).	151
Prolines.	11
*Total **expected** spin systems.*	**140**
- *Missing systems.*	13
- *Unidentified systems.*	7
- ***Unambiguous assignments.***	**120**

Figure 4.14. *A) Graphic summary of the first CRD assignment. Color code: black: not assigned/not reported; grey: prolines; green & red: reported in the BMRB; green: reported in this work as well; red: missed spin systems. B) Assignment statistics. Percentages of the Cα, Cα-1, Cβ and Cβ-1 resonances identified for the two used DC-SIGN CRD samples compared with those of the BMRB file.*

Now, the standard triple resonance protocol successfully provided a good set of $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts that served to certainly assign 120 spin systems. Luckily, almost all the amino acids from the long loop region near to the binding site could be unambiguously assigned, what would be

subsequently helpful for the titration analyses. In addition, some of the discrepancies with the deposited BMRB could be clarified. Fittingly, some mistakes were detected and justified as well (Figure 4.15):

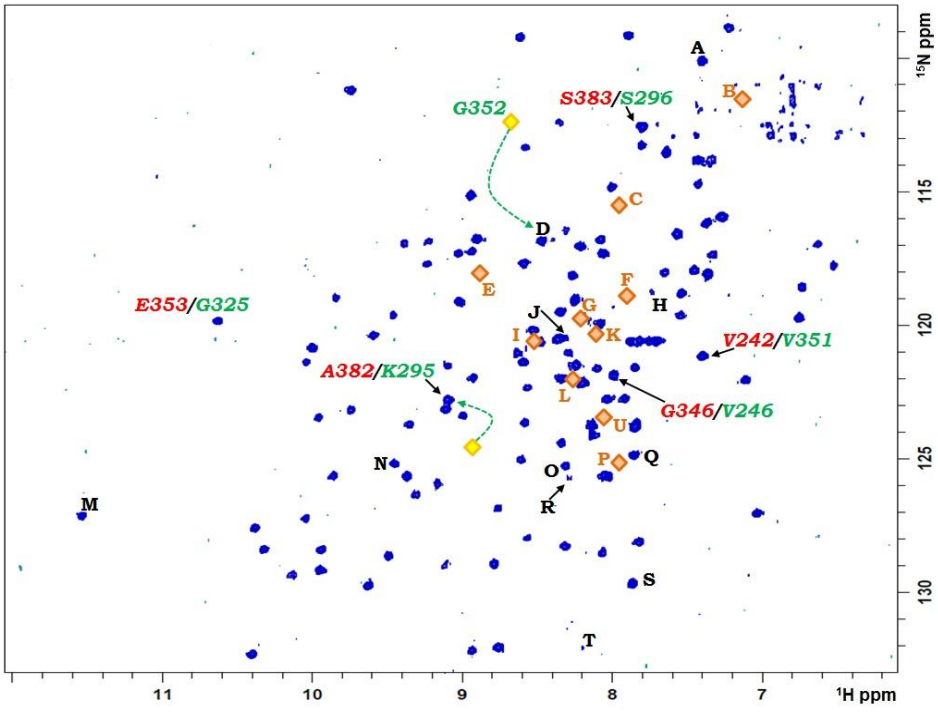


Figure 4.15. *Distribution and analysis of those spin systems whose identification was problematic or controversial. Label color code: black: new crosspeaks not reported in literature; orange diamonds: position of those crosspeaks previously detected for the apo CRD, but vanished for the holo CRD; yellow diamonds: position of G325 and K295 as described in the deposited data (the green arrow indicates the new position found in this work); red: old wrong assignments in the BMRB; green: corrected assignments.*

- After a closer inspection of the ^{1}H-1D spectrum, a weak signal was detected at very low field (peak **M**, 11.53 ppm), but given that all the tryptophan sidechains had been already identified, the spectral window of the ^{15}N-HSQC was enlarged to check whether this signal could be another spin system. Indeed, it actually was, and strikingly, the sequential analysis linked that system to peak **N** and V242. At the same time, V242 was noticed to unambiguously precede the systems **D** and **A**, which was a priori impossible since the next residue type in the sequence is proline (P243).

  Other two connections with unidentified systems were found out (**N-M**-V242-**D-A-Q-O**) and eventually, the peak **O** was discovered to be the amino acid prior to C356. Then, the identified fragment turned out to be

the group of amino acids from the long loop surrounding the Ca^{2+} ion: peaks **N, M, D, A, Q and O** were respectively assigned to **N349, N350, G352, E353, E354 and D355** (Figure 4.16 and Figure 4.17B). To note, G352 had been already assigned but its position in the deposited file was rather shifted at high field in the ^{15}N dimension with respect to our free CRD. Also obvious, the HN pair corresponding to V242 had been wrongly identified and it really belongs to **V351**.

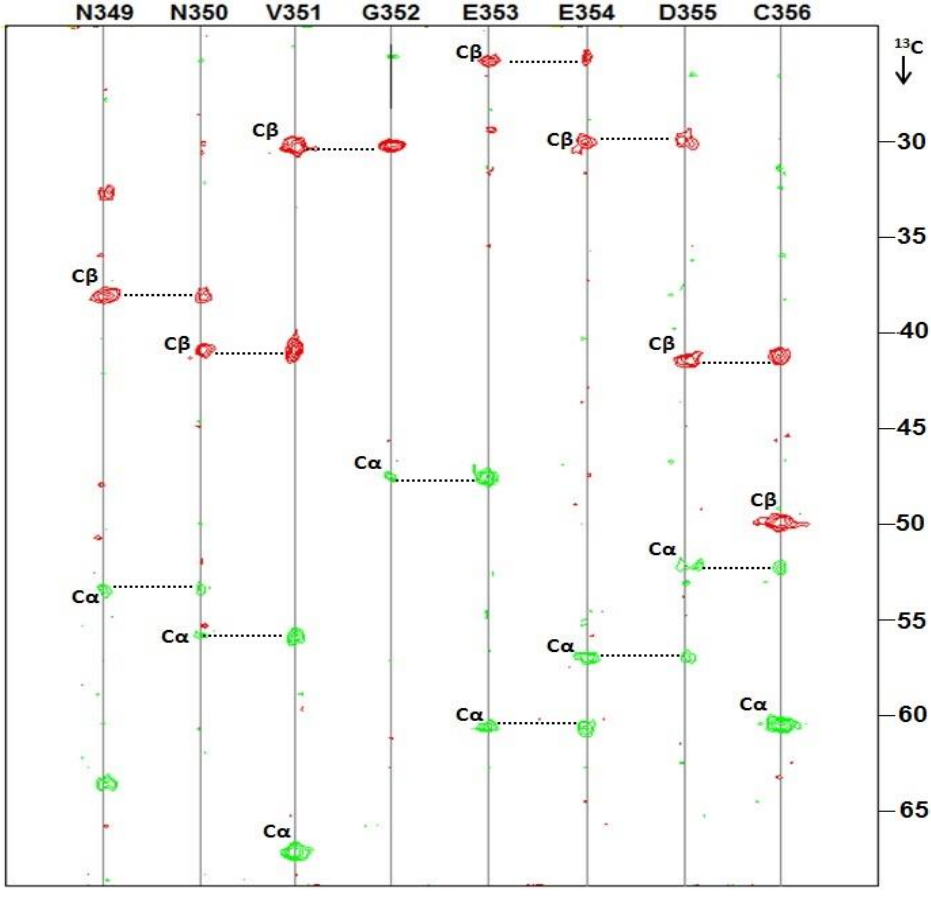


Figure 4.16. *^{13}C-derived sequential correlations established between residues belonging to the loop segment N_{349}NVGEEDC, located at the binding site. Composition made by combination of the carbon strips from experiments HN(CO)CACB and HNCACB.*

- As a result of the previous deductions, other two conflicts with the deposited data were identified: the peak supposed to be E353 was reassigned to **G325,** which was supported by its low Cα chemical shift (45.04 ppm) and the sequential connections with the flanking residues.

The signal of G346, an amino acid from the long loop, displayed an unusual Cα chemical shift, too high for a glycine (59.70 ppm), as well as a clear Cβ correlation. After a thorough analysis of the remaining unassigned pairs, it was assigned to **V246** (His-Tag). Similarly, the sequence A382-S383 did not fit with their expected Cα and Cβ resonances. Instead, it was better identified as the **K295-S296** fragment (Figure 4.17A).

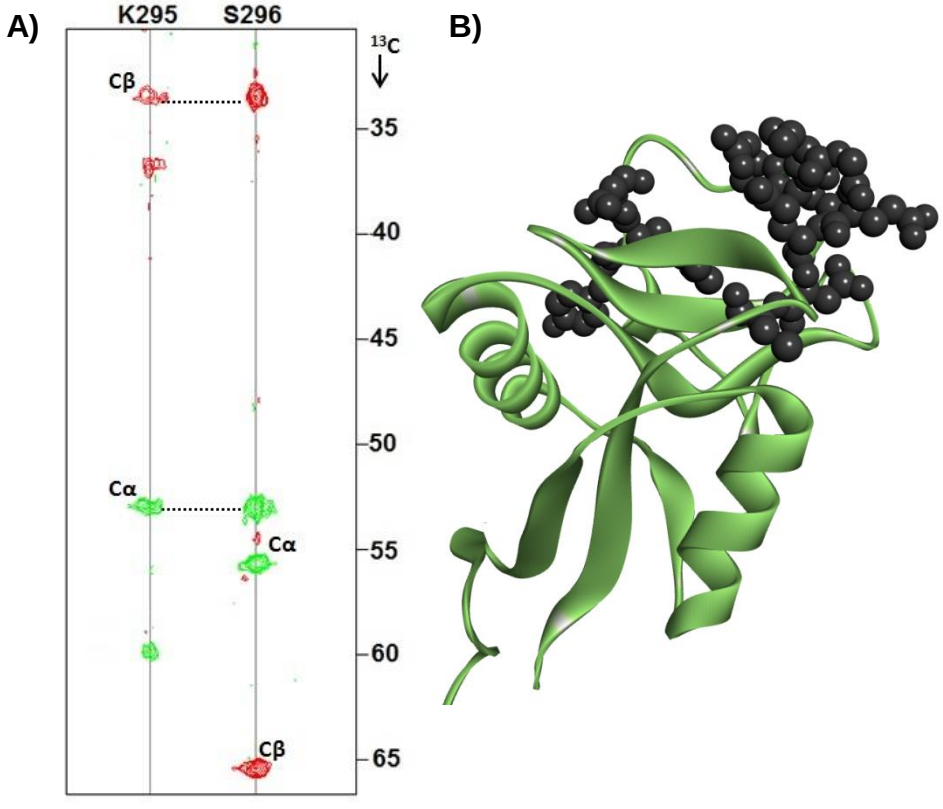


Figure 4.17. *A) Analysis of the ^{13}C resonances extracted from the fragment K295-S296: the strip of K295 shows both Cβ and Cβ-1 ranging around 30-40 ppm, which fits with the pair I294-K295 better than with two consecutive Ala residues (A381-A382). B) Top region of the DC-SIGN CRD in the PDB 1SL5. The black portions represent the regions of the long loop whose assignment was achieved.*

- Regarding the remaining unknown systems, T, R and S were identified as alanines by means of their Cβs resonating at high fields (18-21 ppm). **R** was presumably **A397** due to its closeness to the position registered in the BMRB; **S** was the most intense peak in the 2D spectrum, and meant

to be the C-terminus (**A404**) as its high flexibility surely underlies such a high peak intensity. By discard, peak **T** should correspond to **A395**.

Apart, carbon resonances registered for peak **H** only fitted with **H254.** Most of the remaining peaks were noticed to disappear after adding D-Man to the CRD sample (highlighted as orange diamonds in Figure 3.6.). Some of them were roughly identified as HN pairs already assigned by Prestegard´s group, but marginally shifted (**B→S310, C→T398, E→E298, F→N311, G→Q390, J→E388, K→F391**). The rest were left as unknown spin systems (I, L, P and U). Intriguingly, residues S310-N311 and the aforementioned K295-S296 pair are respectively located at the end and the beginning of the α_2-helix. All of them deeply broadened and/or moved after the addition of D-Man, suggesting fine changes in their dynamics triggered by the presence of the monosaccharide.

Considering all these corrections, 86% of the assignment was completed (Figure 4.18). Excluding the proline residues, 127/140 spin systems were identified in the amide backbone region, and 120 of them unambiguously assigned to specific residues of the CRD sequence. All the chemical shifts were arranged in an STR file and referenced using TSP as NMR standard. The resulting file was submitted to the BMRB database with the accession code **27854** [49].

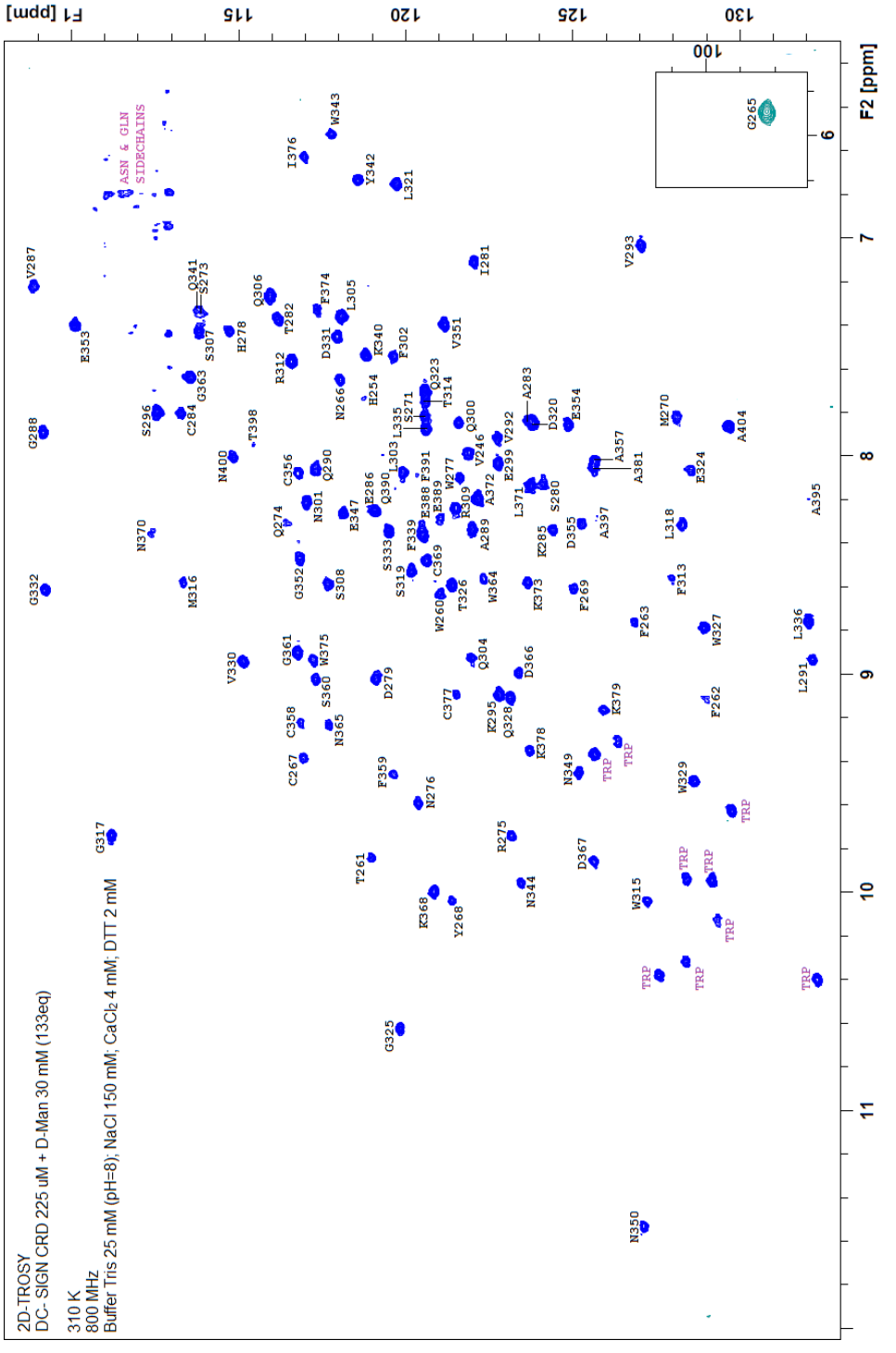


Figure 4.18. *Full CRD assignment performed at 800 MHz, 310 K and in the presence of 133 equivalents of D-Man.*

Taking advantage of the finished assignment, the obtained chemical shifts could be further processed to assess the stability and the integrity of our CRD construct. A good feature to verify that a given protein is correctly folded is to check the existence of intramolecular disulfide bonds. As previously reported, the oxidized or reduced state of the sulfur atom strongly modulates the chemical shift of the nearby Cβ [50]. In the cysteine moiety, this carbon usually displays chemical shifts between 25-30 ppm, whereas the same atom shifts to lower fields when cysteine is oxidized to cystin (Cys-S-S-Cys), showing values within the 38-43 ppm range. For our construct, five out of seven cysteines could be completely assigned, and their Cβ chemical shifts ranged from 36 to 49 ppm (Figure 4.19). Consequently, they all were supposed to take part in the formation of disulfide bonds, forming the pairs C284-C377, C356-C369 and C256-C267. Obviously, although C256 was not identified, its counterpart C267 displayed a much shifted Cβ, thus confirming the presence of a cystin unit. In principle, C384 remained free as reduced cysteine in a flexible region of the CRD, but without discarding the possibility that it could transitorily bind to another C384 from a CRD encountered in solution.

	Cβ chemical shift (ppm)	Expected oxidation state
Cys256	??	Oxidized
Cys267	43.06	Oxidized
Cys284	36.10	Oxidized
Cys356	49.26	Oxidized
Cys369	46.49	Oxidized
Cys377	40.32	Oxidized
Cys384	??	In principle, reduced

Figure 4.19. *CRD DC-SIGN cysteines and their oxidation state deduced from their Cβ chemical shifts.*

Apart, the backbone chemical shifts (H, N, Cα, Cβ) were analyzed by the RCI software [51] and the TALOS software [52] to check the protein flexibility and its secondary structure, respectively. The RCI analysis confirmed the evident high flexibility expected for the N-terminal and C-terminal regions. Moreover, another region with notable motion corresponded to the short loop segment situated behind the long loop and the primary binding site (Figure 4.20). Surprisingly, the long loop seemed to preserve a marked rigidity, which was expected to be caused by the presence of D-Man, and would confirm again that ligand recognition actually increases the structural stability of this loop.

In parallel, the secondary structure elements predicted by the TALOS+ tool fitted quite well with those identified in the deposited X-Ray crystals (Figure 4.21). The unique discrepancies were detected for the β1' and the β2' strands: the registered chemical shifts pointed out that these regions could be more ordered than expected, compared with the previous existing data. However, the β2' segment is located on the back part of the CRD, and the TALOS+ analysis would place the extended β2' into a hinge region which establishes contacts with the calcium ions at sites 2 and 3. It cannot be ensured whether this is an artifact due to miscalculations or it is actually as rigid as predicted.

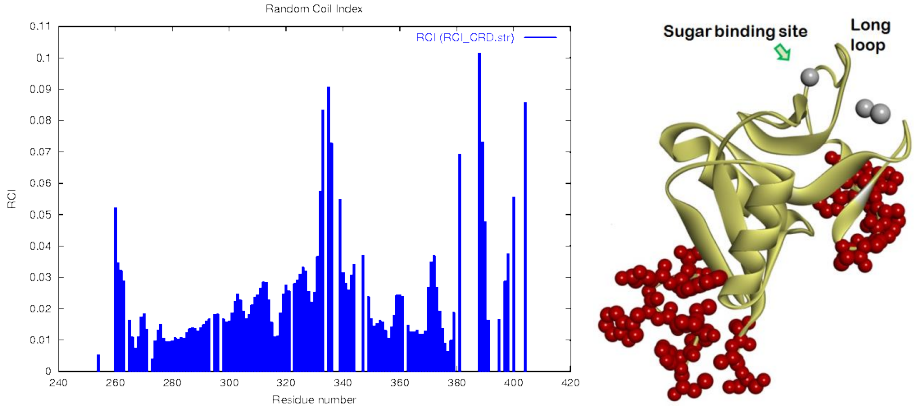


Figure 4.20. *On the left, the Random Coil Index (RCI) based on the H, N, Cα and Cβ chemical shifts of the assigned residues. On the right, the CRD of DC-SIGN showing those residues with the highest RCI values as CPK spheres.*

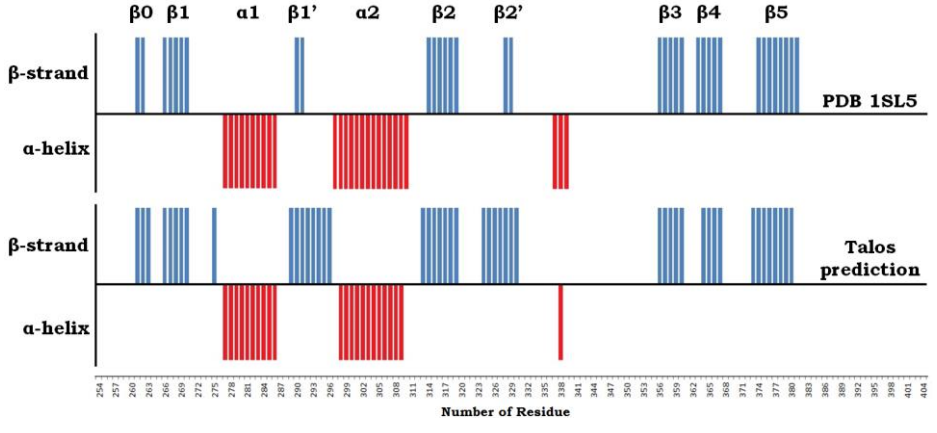


Figure 4.21. *Expected secondary structure for the CRD based on the backbone chemical shifts. Comparison between predictions made by TALOS+ and the secondary elements present in the PDB 1SL5.*

4.3. HSQC-based tritrations

To analyze the DC-SIGN CRD perturbations upon ligand binding, successive ^{15}N-HSQC spectra were recorded at increasing amounts of the oligosaccharides of interest. As explained in the *Methods* section, the working data set is constituted by the ^{1}H and ^{15}N backbone resonances of the protein, one from each amino acid residue, originated by their corresponding α-amino groups. The resulting 2D spectrum is called the protein "fingerprint", as the correlation pattern distribution is characteristic and unique for each protein or peptide.

^{15}N-HSQC-based titrations are suitable [53-54] for describing binding processes with affinities that are weak but still relevant in the biological context (from 10 µM to 10 mM). Usually, the upper limit of this technique is set around 10 mM, what includes the system studied herein. In fact, ^{15}N-HSQC-based titrations have been already performed by other research groups to study related lectins, like Langerin [55] or L-SIGN [32], as well as DC-SIGN [38,56].

To collect a good set of experimental data, the peak intensities and the spectrum resolution were maximized by recording the titration points at 800 MHz, in a spectrometer equipped with a cryo-probe, and using the TROSY pulse sequence for the acquisition of the ^{15}N-HSQC spectra.

Firstly, the binding site on the lectin CRD surface was mapped for each tetrasaccharide independently, making use of the same ^{15}N-CRD batch with a concentration of 120 µM. Apart from the two blood group antigens, L-fucose (Fuc) was included in the analyses to account for possible differences in the binding processes arising from the surrounding sugars. Noteworthy, from the equilibrium constant, it can be easily deduced that the total ligand concentration at the middle point of the titration is almost equivalent to the dissociation constant as long as $[L_0]>>>[P_0]$. At this point, the half of the total amount of protein is forming the complex ([P]=[PL]), while the ligand concentration can be approximated to $[L_0]$:

$$K_d = \frac{[P][L_0]}{[PL]} \xrightarrow{At\ 50\%\ sat.} [P] = [PL] \rightarrow \mathbf{K_d} \approx [\mathbf{L_0}] \qquad (Eq.1)$$

Taking into account the low affinities expected for Fuc and related antigens [36,57], the middle points of the titrations were expected at ligand concentrations ranging around 6 mM (Fuc), what implies using a

concentration of about 50-60 equivalents with regard to the ^{15}N-CRD. To achieve almost complete protein saturation and observe the maximum chemical shift perturbations ($\Delta\delta_{max}$), sequential ^{15}N-HSQC experiments were acquired until reaching 140 eq. of ligand (16.8 mM). The data were processed by employing Eq. 8 described in the *Methods* section, with $\alpha = 0.14$ [58]. The resulting CSPs are summarized for Fuc and the A and B blood groups in Figure 4.23.

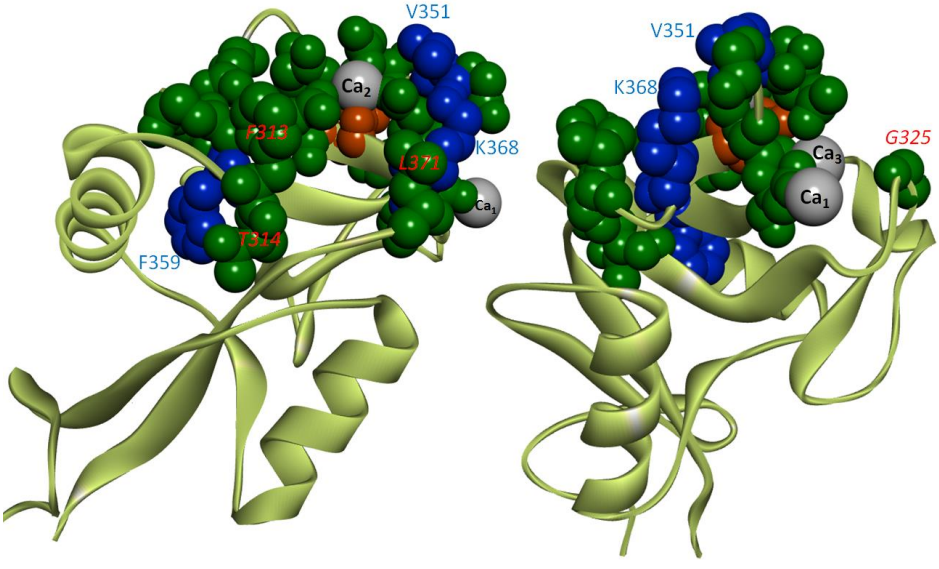


Figure 4.22. *Front and side views of the DC-SIGN CRD showing those amino acids which yielded the strongest CSPs. In green, the residues shifted the most upon titration with the B antigen (16 AA). In blue, the three additional residues which were notably shifted by the A antigen as well. In orange, D366 (very shifted for the B antigen and not observed for the A antigen). Red labels indicate those perturbed amino acids placed away from the binding site.*

Comparing both tetrasaccharides, few differences could be detected, as they displayed very similar CSP patterns. In both cases, the most affected amino acids were those directly contacting the calcium ion at the primary binding site, belonging either to the top part of the flexible long loop or to the β_4-strand that goes in parallel. Some residues located on the hinge between strands β_3 and β_4 were likewise shifted (S360, G361, G363). On the long loop, perturbations were notably propagated from N344 to the beginning of the β_3-strand (C356), including few residues which also established coordination bonds with the calcium ions at sites 2 and 3. Immediately, the localized distribution of the most intense CSPs allowed confirming that both binding processes are taking place in a specific fashion, by chelating the metal ion at

site 1 (Figure 4.22). Interestingly, there were several amino acids which lay further away from the most affected area. Some of them, like F313, T314 or L371, are placed into the secondary binding surface defined by strands β_2, β_3 and β_4 where other sugar residues have been described to establish stabilizing VdW interactions [59-60]. On one hand, these findings suggested that both ligands can make additional contacts through other sugar residues, although this cannot be fully demonstrated by receptor-based methods. On the other hand, CSPs at remote positions can simply come from the primary perturbations triggering long-range effects propagated through the network of H-bonds [61]. The latter case could explain certain long-range CSPs detected in other parts of the CRD construct, like those observed for G325 and D320.

In absolute terms, the collected CSPs were slightly higher for the B antigen than for the A antigen, and there was a notorious difference between them in the data dispersion assessed by the standard deviation (Figure 4.24A). Even so, the blood group A showed marked perturbations at residues F359, K368 and V351 (Figure 4.22), not noticed for the B antigen. In constrast, E353 was largely less affected by the presence of the blood group A, although the observation of this residue was rather difficult due to the intense signal overlap, and an accurate measure of the CSP was impossible. Surprisingly, the CSP pattern obtained after titrating the CRD with L-Fuc was very similar to those reported for the blood groups. Moreover, L-Fuc seemed to trigger larger CSPs at V351, E353 and D367, all of them located on the same side of the binding site (Figure 4.24B).

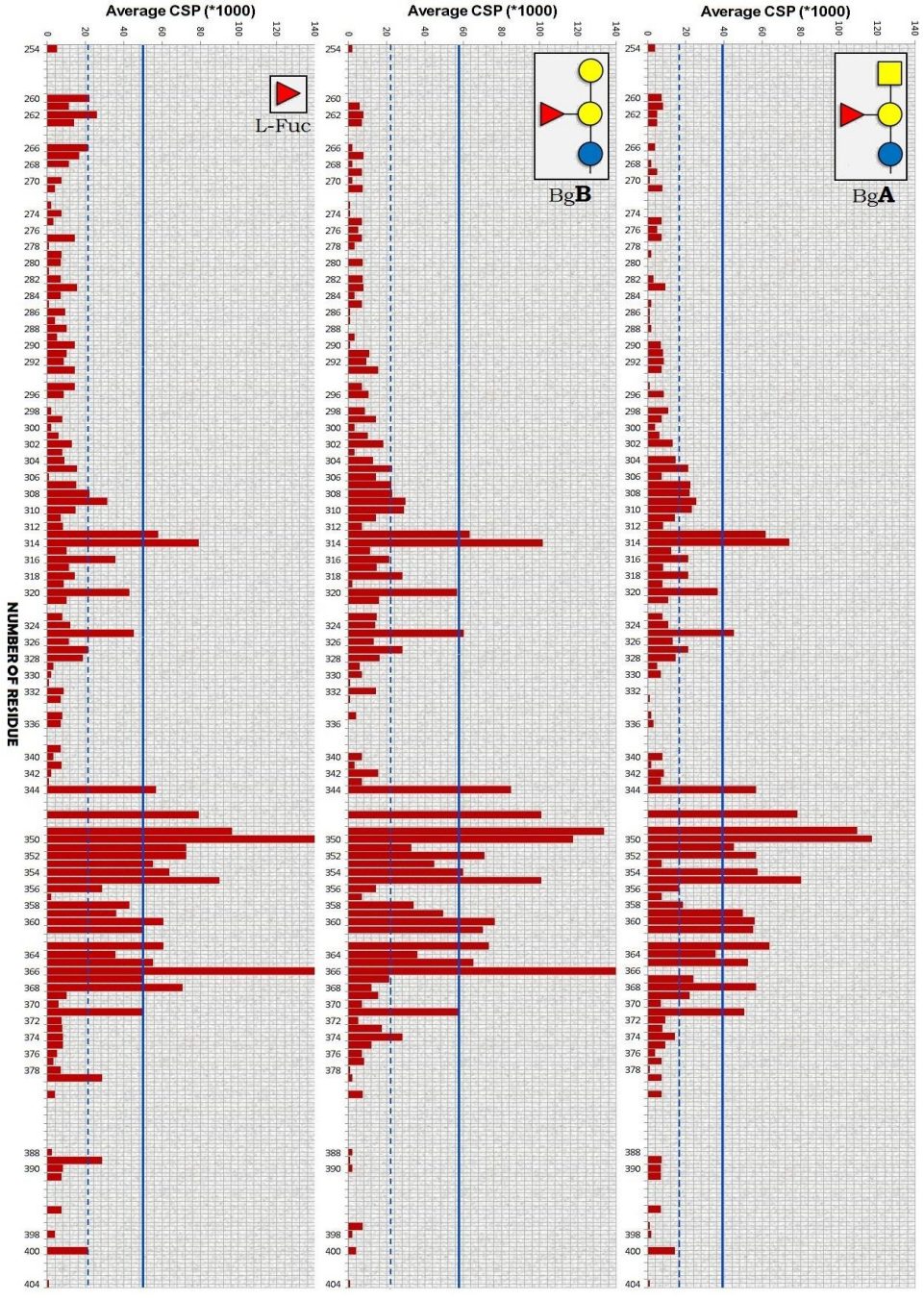


Figure 4.23. *CSPs for the Fuc-containing fragments measured in ^{15}N-CRD samples at 120 µM containing 140 eq. of each ligand respectively (310 K).*

A)

	Mean (μ)	Stdev. (σ)	Cutoff (μ+σ)	CSPs>Cutoff
Blood Group A	16.6	23.0	39.6	**18**
Blood Group B	21.7	35.4	57.2	**16**
L-Fucose	20.9	28.8	49.7	**19**
D-Man	18.9	22.1	41.0	**14**
Man_3	19.6	33.7	53.2	**15**

B)

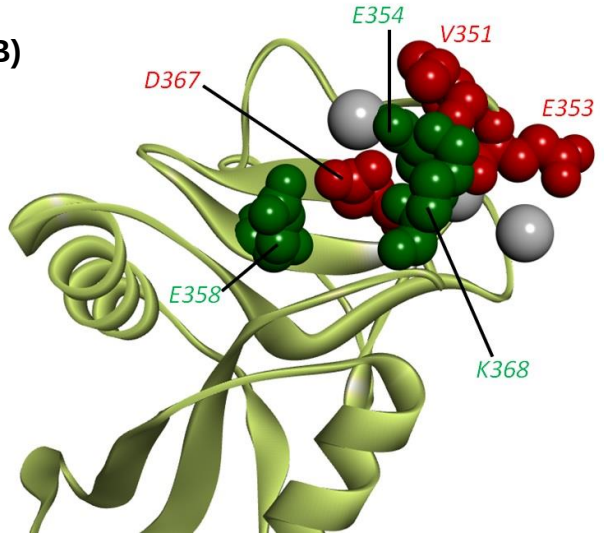


Figure 4.24. *A) Statistical analysis of the data collected from ^{15}N-HSQC spectra for the three Fuc-containing and the two Man-containing ligands. The values are expressed in CSP*1000 (ppm). B) Most meaningful differences noticed during CSP data analysis. In red, those residues more shifted by Fuc than by the antigens A and B. In green, some amino acids notably perturbed by Man and Man_3 that remained less affected in the presence of Fuc-containing fragments.*

In parallel, other two ^{15}N-HSQC titrations were carried out in the presence of the Man epitope instead of Fuc. Man and Man_3 (core trimannoside) were tested and compared with the fucosylated fragments (Figure 4.26). At first sight, Fuc and Man seemed to share very similar CSP patterns, although Man displayed an overall reduced effect on the protein chemical shifts. The only significant difference between Man-containing and Fuc-containing ligands was the particularly strong CSP caused in K368 by the presence of Man. There were other minor differences regarding nearby residues in the β-sheet like (E358, N365) and the long loop (E354, D355) (Figure 4.24). However, the intriguing point here arose when the monosaccharides were compared with their respective related oligosugars: Man and Man_3 caused analogous perturbations on the CRD, as Fuc and both blood groups did as well. The fact that Fuc and Man likewise contributed to the existence of long-range CSPs

suggested that all the reported perturbations are exclusively triggered by the residue that binds to Ca^{2+}. Consequently, it should be thought that the surrounding sugars do not establish tight stabilizing interactions with additional amino acid residues exposed on the CRD surface, but they would presumably play minor roles in the slightly improved affinities reported for the oligosaccharides. This hypothesis was reasonably supported by the already existing crystal structures of DC-SIGN CRD in complex with Man_4 and Man_6, both fragments containing the Man_3 core [59-60]. In both models, the Man at the primary binding site targets the closest amino acid sidechains by means of multiple H-bonds through its OH3, OH4 and OH6 groups. In contrast, there are no more H-bonds arising from the other Man residues in spite of lying very close to the surface, except for an isolated H-bond established by the α1-6 linked Man (Figure 4.25). In principle, the same would be expected to happen in the case of the blood group antigens, but additional information is needed to accurately delineate the binding mode and explain the possible lack of VdW contacts through sugar residues flanking the Fuc moiety.

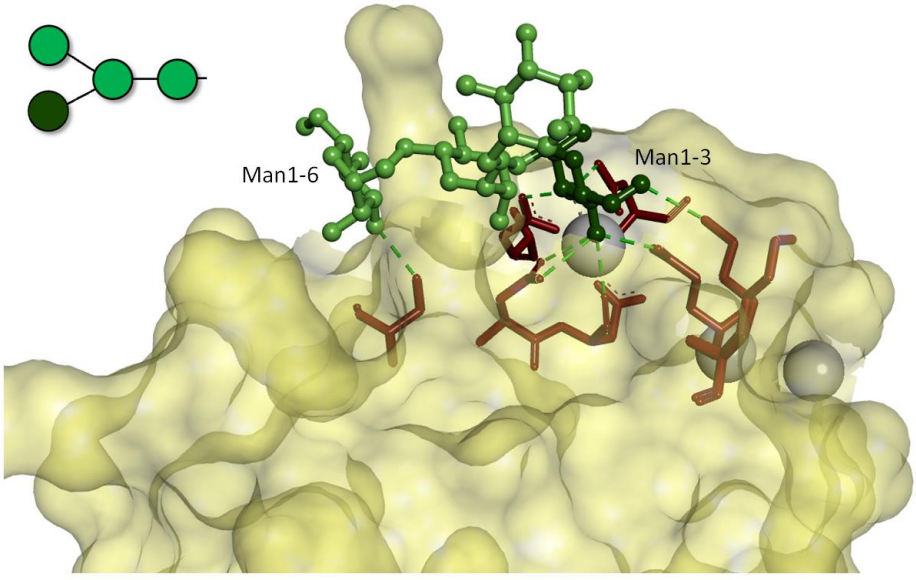


Figure 4.25. *DC-SIGN CRD-Man_4 complex (PDB: 1SL4). The binding Man is highlighted in dark green. The amino acids involved in VdW contacts with the ligand are shown as stick bars, and H-bonds are indicated by dashed light green lines. White spheres represent the calcium ions.*

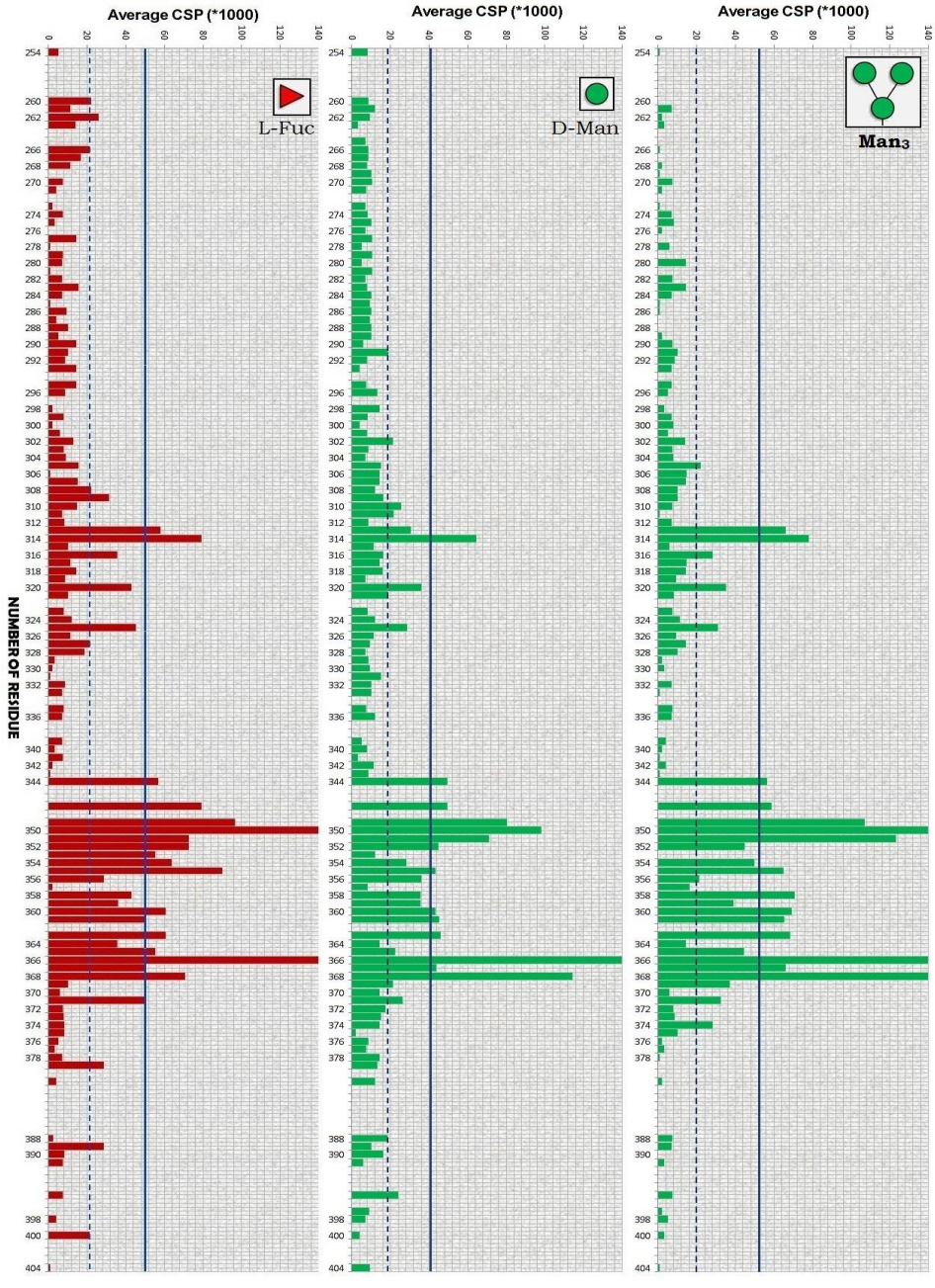


Figure 4.26. *Comparison between the CSPs collected in the presence of L-Fuc and after adding D-Man and the Man-containing core trimannoside (Man_3).*

Apart from mapping the binding site, the dynamic behavior of the CRD was investigated and the affinities were estimated by means of building the corresponding binding isotherms. As mentioned above, the middle points of the titration are reached when the ligand concentration approaches to the K_d value. To minimize the experimental errors during the curve fitting, it is advisable that many points are acquired within this range, using ligand concentrations close to the theorical K_d [62]. In total, ten points were recorded for each ligand, and given that the weakest sugar was expected to be Fuc, most of them corresponded to protein/ligand ratios below/around 1:60 equivalents.

Firstly, the titration spectra were scrutinized to analyze the type of exchange regime. Interestingly, for both tetrasaccharides many residues were noticed to remarkably broaden immediately after adding the ligand (10-20 eq). In fact, N349, N350 and D366 cross peaks were almost lost during the middle points of the titration (Figure 4.27 and 4.28). In the case of the blood group B, the peak intensity could be subsequently recovered at high ligand ratios, even exceeding in some cases the initial intensities corresponding to the free state, as observed for N350 or G325. This sharp broadening process followed by signal recovering could be reported for almost any shifted peak, either belonging to the primary binding site or placed further away (F313, G325). Clearly, the observed behavior was indicative that an **intermediate exchange process** takes place upon binding of the B antigen (Figure 4.27). For the blood group A, a slightly different scenario was found: in general, signal broadening affected to many shifted peaks, but intensities were marginally or not recovered when the ligand concentration increased. After initial intensity decay, the broadening was maintained without progressing or reversing. For some residues located at the long loop, as N349, N350 or D355, a slight signal recovery was detected as well, thus suggesting the same intermediate exchange regime as that documented for the B antigen. In contrast, other amino acids showed no relevant changes in the broadening along their evolution (G325, N365), and their cleaner drifts better suggested the existence of a fast exchange regime (Figure 4.27).

In either case, the peak displacements always followed linear trajectories, and therefore the possible binding of two or more ligand molecules to the DC-SIGN CRD was discarded. Nevertheless, the existence of non-specific binding could not be ignored, although its effect in CSPs is usually negligible since these transitory interactions at the receptor surface are extremely weak [63]. For the blood group B, there was a clear intermediate exchange derived from

an exchange rate very close to the experimentally measured CSP values (in Hz), while the binding of the blood group A seemed to take place in a transient regime ranging from intermediate to fast exchange. Considering that both tetrasaccarides are almost structurally identical in terms of molecular size and diffusion, binding processes were expected to take place under the same k_{on}, and therefore, these results would point out that the B antigen exhibits a slower k_{off} than the A antigen, what means that the former would have a slightly better affinity for the CRD.

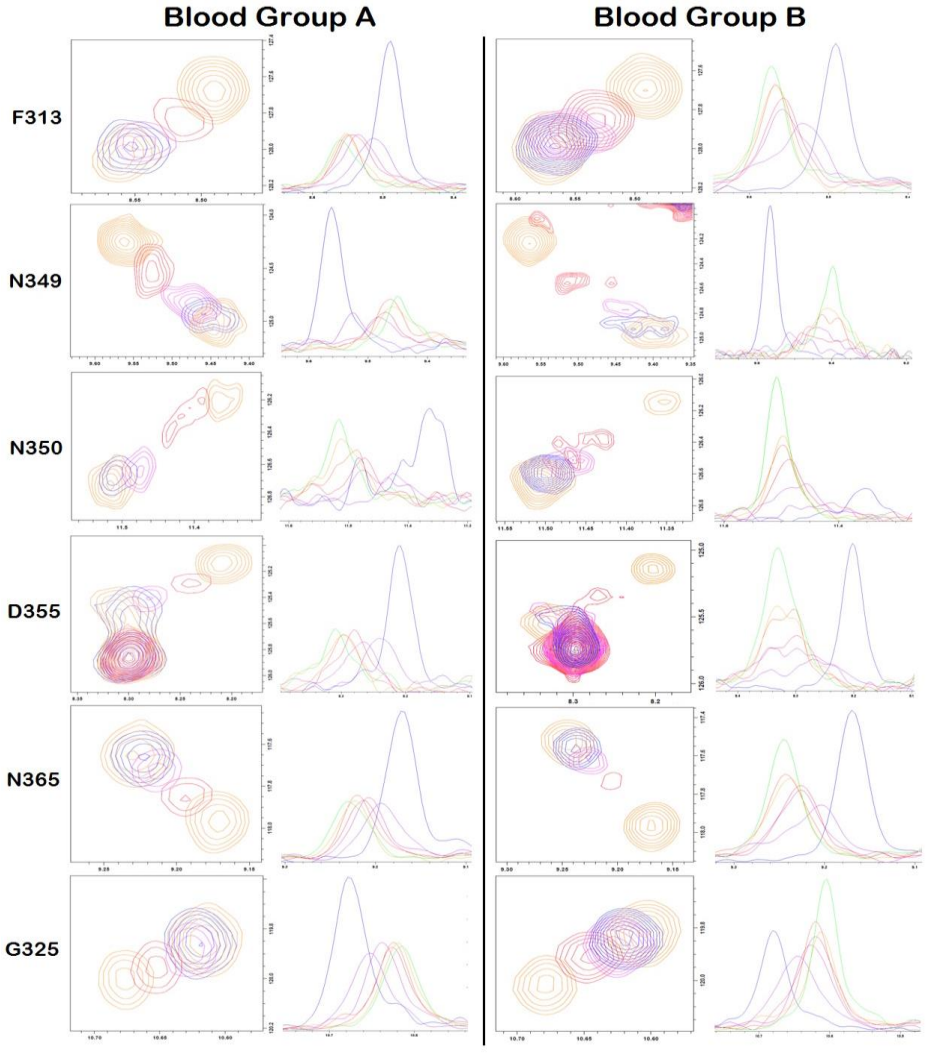


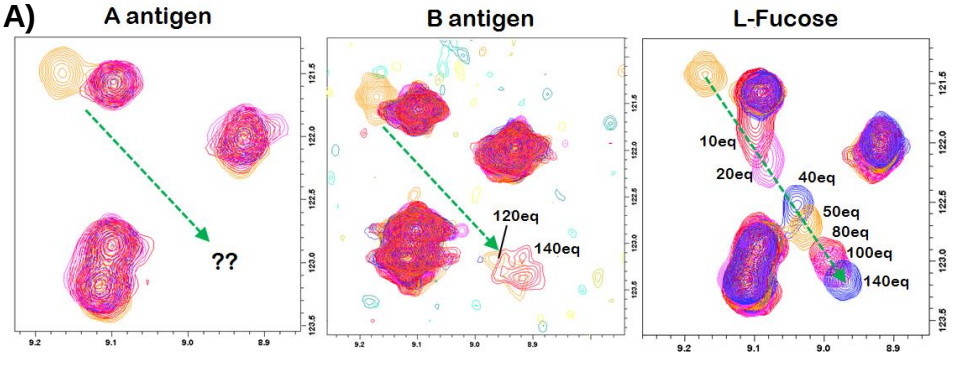
Figure 4.27. *Crosspeak evolution reported for six representative amino acids. The box on the left shows the observed shifts in the 2D-HSQC (orange: free CRD; red: 20eq; pink: 60eq; dark blue: 100eq; orange: 140eq), and the picture on the right shows the 1D profile to better compare peak intensities (blue: free CRD; purple: 20eq; pink: 40eq; red: 60eq; orange: 80 eq; pale orange: 100 eq; green: 140eq).*

In accordance with reported data [64], the existence of progressive signal broadening can be related to the presence of a higher number of receptor conformers derived from ligand recognition, although other feasible reasons can also explain this phenomenon, including multiple or non-specific binding [53]. However, and specially for the blood group A, it is unlikely that these processes were actually occurring, as multiple binding events usually lead to exaggerate and crescent signal broadening, eventually causing that correlation peaks vanish at high ligand concentrations. Instead, for the A antigen, signal broadening remained almost constant even in the presence of very large amounts of ligand. This behavior is more likely to arise from subtle conformational changes triggered into receptor´s structure and/or from complex effects of the fast-intermediate regime itself. To note, weird line shapes were also noticed for few residues alongside extreme broadening, sometimes giving rise to very wide signals with irregular patterns, which probably come from protein conformers that coexist at low protein-ligand ratios (N350, D355) and finally converge to an unique and stable conformer (bound receptor) when the ligand amount increases. At the same time, as mentioned before, some correlation peaks became sharper signals at the saturation point when titrated with the B antigen, what suggested that the area around the primary calcium ion (specially the loop regions) gains stability upon sugar attachment.

Experimentally, the exchange rates could be roughly estimated by using the titration data from residues N349 and N350. D366 was also a good candidate, since it was the most prone to reach coalescence. However, it disappeared in most of the recorded points for the titration with the blood group B, and in fact, the intense line broadening provoked by binding of the A antigen precluded at all the observation of that residue at the saturation point (Figure 4.28A). Applying the Eq. 7 described on the *Methods* section, the exchange rate k_{ex} was estimated around 100 Hz (Figure 4.28B). As reported, the kinetic constant k_{off} commonly takes values between 10^2 and 10^4 s^{-1} for processes occurring within an intermediate exchange regime [65]. This means that the product $k_{on}[L]$ must never surpass two orders of magnitude to fulfill the equivalence stated in Eq. 7 for k_{ex}. Given that the ligand concentration around the coalescence point is about 10-15 eq (~1-2 mM), the expected k_{on} would stand within the 10^3-10^5 M^{-1} s^{-1} range, what suitably fits with reported kinetic constants for other carbohydrate-binding lectins [66-68]. Also, this estimation underlines the otherwise false assumption on supposing a diffusion-driven recognition process and considering a k_{on} of about 10^8 M^{-1} s^{-1} [69]. Certainly, the estimations done highlighted that the binding of both antigens is very

similar from a kinetic point of view. Indeed, the on and off rate constants were within the expected range of values in each case. The k_{on} rates took values around 10^4 M^{-1} s^{-1}, although the binding process seemed to be slightly faster for the B antigen. There are many factors that may account for slowing down the on rates under the diffusion control limit. Often, productive collisions between receptor and counterpart are hampered by processes of conformational selection [70]. According with our gathered data so far, it is likely that the conformational flexibility of the binding site region on the DC-SIGN CRD is behind these k_{on} values. Likewise, both k_{off} values perfectly fitted with an intermediate exchange regime, ranging around 10^2 s^{-1}. It is worth mentioning that the calculated k_{on} and k_{off} were very approximate, and additionally, the coalescence is probably reached below 10 eq, although the exact position could not be determined with enough accuracy. Then, it was difficult to conclude where the difference in affinity between both tetrasaccharides is arising from, although the most probable reason might be the magnitude of their on rates rather than the k_{off}, as thought at first.

A)



B)

$$\mathbf{k_{ex}} = \frac{\boldsymbol{\pi} \cdot \boldsymbol{\Delta\omega}}{\sqrt{\mathbf{2}}} = \boldsymbol{k_{on}[L] + k_{off}}$$

	N349			N350		
	Δω (Hz)	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})	Δω (Hz)	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})
10eq BgA ***K_d ≈ 7.1mM***	87.68	**$2.3 \cdot 10^4$**	**$1.7 \cdot 10^2$**	93.94	**$2.5 \cdot 10^4$**	**$1.8 \cdot 10^2$**
10eq BgB ***K_d ≈ 2.3mM***	106.97	**$6.8 \cdot 10^4$**	**$1.6 \cdot 10^2$**	93.97	**$6.0 \cdot 10^4$**	**$1.4 \cdot 10^2$**

Figure 4.28. *A) Different CSP patterns reported for D366 in the presence of both tetrasaccharides and L-Fuc. B) Estimation of the binding kinetics through CSP from N349 and N350, assuming that peak coalescence is achieved at 10 eq for both sugars (1.2 mM). As can be noted, the off rates are very close in all cases whereas the k_{on} values seem to be slightly higher for the B antigen.*

Finally, the binding isotherms were calculated by adjusting the titration data to Eq. 10 described in *Methods*. Very interestingly, a notable difference was noticed between both oligosaccharides when plotting the evolution of the CSP as a function of the ligand concentration. The blood group A yielded normal isotherms in all cases, as expected for a simple binding event, with a 1:1 stoichiometry (Figure 4.29). However, a completely different scenario was found out for the B antigen: the data did not adjust to a unique crescent curve, but it reached an intermediate plateau around 40-50 eq of tetrasaccharide, and grew again until protein saturation (120-140 eq). Noteworthy, it was impossible to detect this marked difference exclusively through the visualization of the 2D-HSQC spectra due to the low magnitude of the collected CSPs, but it became clear when representing the data in separated graphs. In principle, this kind of staggered curve is typically observed for receptors with several binding sites that are occupied in a stepwise fashion [71-72]. However, the essentially linear displacement observed for every residue in both cases unequivocally discards the binding to other regions in the CRD surface. Thus, the only plausible possibility was to consider an additional recognition process governed by a weaker K_d, but this minor binding mode would be taking place at the same primary binding site. In summary, there would presumably be a major binding mode for both tetrasaccharides, while only the B antigen would be able to target the primary Ca^{2+} site through two binding modes co-existing at high protein-ligand ratios. In addition, the existence of two modes of interaction could explain the higher k_{on} estimated for the B antigen (better rebinding).

According to previous deductions, the estimated affinities agreed that the B antigen (K_d = 2.3 ± 0.6 mM) is a better binder than the A antigen (K_d = 7.1 ± 1.4 mM) (Figure 4.31). Surprisingly, Fuc turned out to have an intermediate affinity compared with both tetrasaccharides (K_d = 3.7 ± 0.5 mM). In any case, the three affinities were of the same order (mM), as expected for this kind of fucosylated oligosaccharides. The better affinity of the B antigen would be explained by the aforementioned existence of two binding modes (major and minor), whereas there was not an immediate explanation about why the A antigen displays less affinity than Fuc. Certainly, the titrations with Fuc were performed by using the mixture of anomers (α/β) in which all hydroxyls groups are free. As reported both for Man and DC-SIGN [60,73] and for Fuc with other lectins [74-75], the receptor attachment to simple monosaccharides with many sterically available hydroxyl groups may easily take place through multiple binding modes. In fact, the binding isotherms obtained for Fuc resembled, in many cases, the staggered shapes noticed for

the B antigen. Moreover, the evolution of some peaks did not seem to follow a linear trend, but a marked change in their CSP trajectories was not observed either (Figure 4.30). These smooth deviations from linearity could arise from little differences in how each binding mode modifies the CRD chemical shifts, but in no case it can be said that the sugar is bound to another site on the CRD surface. Therefore, the existence of several possible binding modes for Fuc could perfectly account for its lower dissociation constant compared with blood group A.

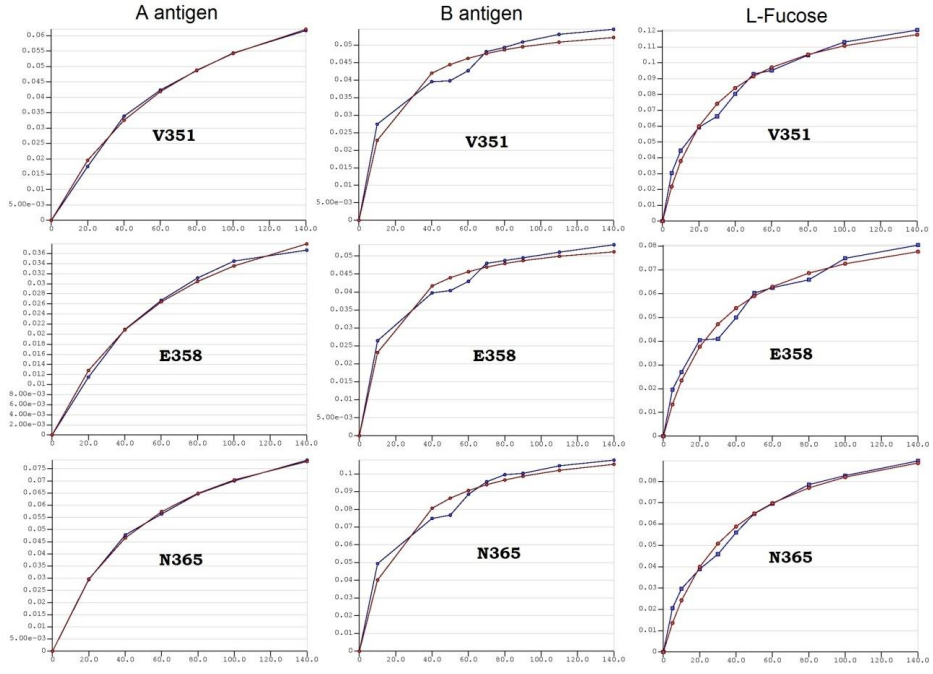


Figure 4.29. *Binding isotherms built by using CSP data from residues V351, E358 and N365. The blue lines represent the experimental data and the red lines correspond to the theoretical curve adjusted by means of the CCPNMR software [76].*

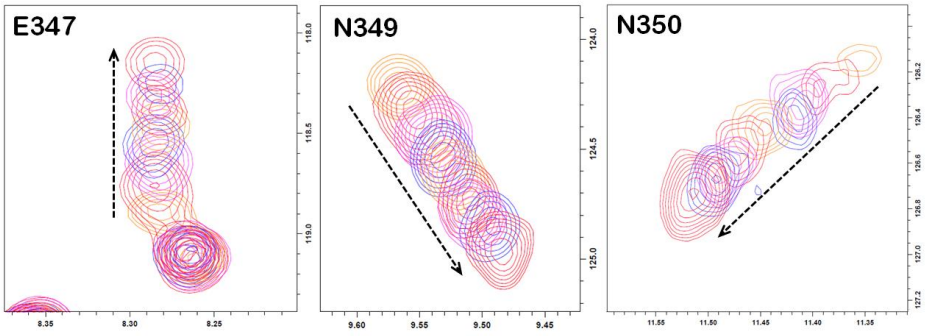


Figure 4.30. *Chemical shift perturbations monitored for residues E347, N349 and N350 along the titration of the DC-SIGN CRD with L-Fucose. In contrast to the blood group antigens, there was not a deep change in the signal broadening for any peak, what suggested that the interaction of Fuc takes place in a fast exchange regime and hence it will be expected to have faster kinetic constants (k_{on} & k_{off}).*

	Blood group A K_d (mM)	Blood group B K_d (mM)	L-Fucose K_d (mM)
Met316	7.1 ± 0.9	2.0 ± 0.5	3.1 ± 0.5
Gly325	5.4 ± 0.6	2.3 ± 0.5	3.3 ± 0.3
Glu347	7.0 ± 0.1	3.4 ± 0.5	3.5 ± 0.3
Asn349	5.2 ± 0.9	-	-
Val351	9.0 ± 0.8	1.7 ± 0.4	3.1 ± 0.2
Gly352	8.1 ± 0.8	3.4 ± 0.2	3.7 ± 0.3
Glu354	10.0 ± 1.0	2.1 ± 0.5	4.6 ± 0.6
Asp355	7.2 ± 0.4	-	3.8 ± 0.2
Glu358	7.8 ± 1.0	1.6 ± 0.3	3.5 ± 0.5
Phe359	5.4 ± 0.4	2.2 ± 0.3	-
Asn365	5.8 ± 0.3	2.2 ± 0.5	4.1 ± 0.1
Lys368	7.7 ± 0.7	-	3.4 ± 0.3
Trp364	7.5 ± 0.6	2.0 ± 0.4	4.3 ± 0.4
Leu371	5.7 ± 0.3	2.9 ± 0.7	3.3 ± 0.4
Average	**7.1 ± 1.4**	**2.3 ± 0.6**	**3.7 ± 0.5**

Figure 4.31. *Estimation of the binding affinities for Fuc and the antigens A and B. The equation applied was $y = A(B+x-[(B+x)^2-4x]^{0.5})$, with $A = \Delta\delta_{max}/2[P_0]$ and $B = [P_0] + K_d$. In the case of the B antigen, K368 was not used as its CSP was negligible, and the extreme broadening affecting N349 and D355 shifts precluded the accurate determination of their CSPs. For Fuc, N349 and F359 were not either employed since the curve fitting could not be suitably done due to very large errors in the estimated parameters.*

4.4. STD and tr-NOESY

The results derived from the HSQC-based titrations essentially pointed out that both tetrasaccharides were specifically recognized through its Fuc moiety and there might be a minor contribution from the surrounding sugars to the binding. Interestingly, different affinities and binding isotherms for both ligands were unexpectedly noticed at the same time. As mentioned in the previous section, the crystal structures of DC-SIGN harboring Man-containing fragments (2IT5 (Man_6), 2IT6 (Man_2), 1SL4 (Man_4), Refs. [59-60]) always show the Man residue targeting the Ca^{2+} ion establishing the major contacts, mainly H-bonds with the neighboring polar sidechains. The only available X-Ray model of DC-SIGN interacting with a Fuc-containing fragment is the PDB 1SL5 [59], which contains the CRD bound to Lacto-N-fucopentaose type III (LNFP III) (Figure 4.32).

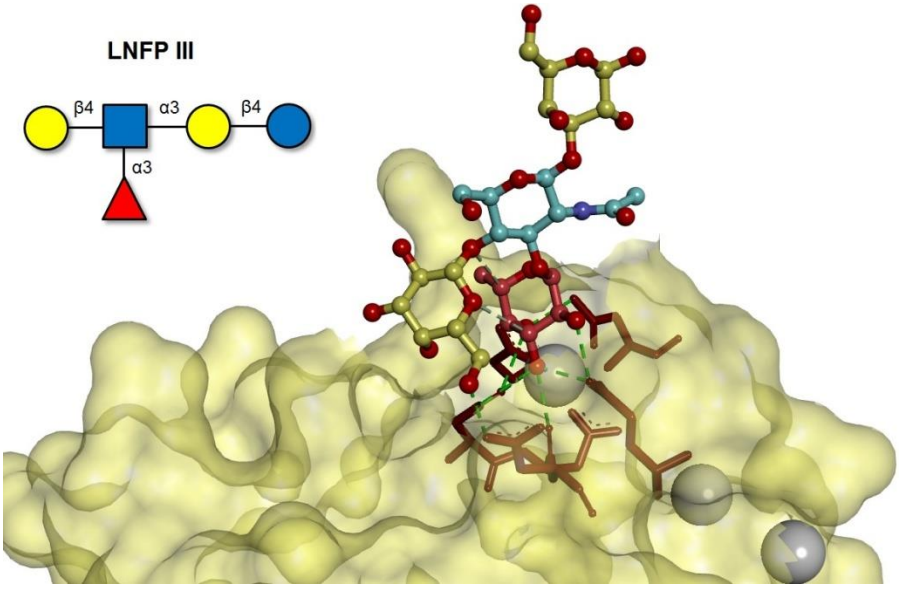


Figure 4.32. *DC-SIGN-LNFP III complex (PDB 1SL5). The reducing Glc is not modeled. The sidechains of the CRD involved in VdW contacts with the ligand (green dashed lines) are detailed as brown sticks. As noted, all the H-bonds are established through the Fuc residue, except for a single H-bond that Gal forms with D367.*

Both NMR and crystallographic data support the proximity of the non-reducing Gal moiety to the β-sheet plateau next to the metal ion, although the X-Ray structure does not display any direct Gal-CRD contact, except for an isolated H-bond between OH6 and D367. Certainly, the NMR models have suggested that the intermolecular distances are likely shorter than those described by X-Ray, but in either case the presence of close surrounding

sugars seems to play an additional role in the recognition event, although it has not been well clarified yet. In this scenario, the results reported herein for the blood antigens are actually fairly interesting since clear differences have been noticed between them. Intriguingly, although the interaction is exclusively driven by the Fuc moiety, each antigen is somehow triggering a distinct dynamic behavior which affects their affinities as well. Obviously, these findings have immediately raised the question about whether those discrepancies could be related to the only structural element that differentiates both blood groups: the non-reducing Gal/GalNAc moiety.

To shed light into this question, the ligand-based approach was put in practice. Namely, additional analyses were conducted by means of STD and tr-NOESY experiments. The primordial aim was to gather enough data to propose a suitable 3D-model which would be subsequently evaluated by computational simulations.

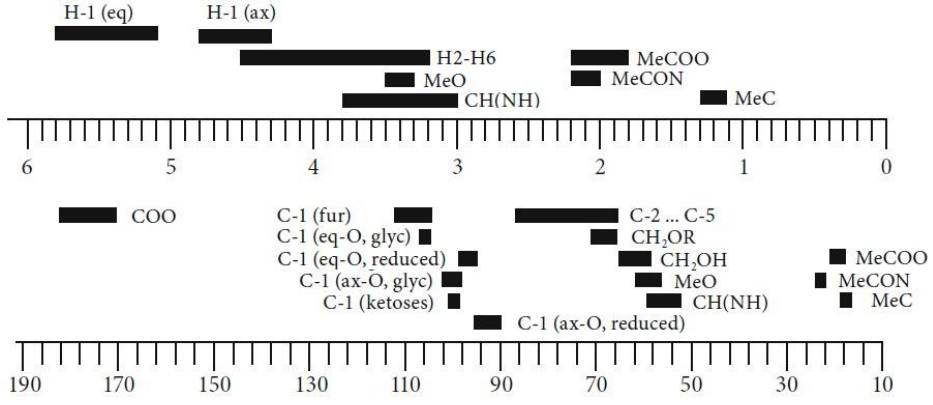


Figure 4.33. *Typical ^{1}H and ^{13}C chemical shift ranges found in carbohydrates. Adapted from [77].*

Previously, the NMR chemical shifts of the two tetrasaccharides had to be assigned. Only isolated monosaccharides can be readily identified in solution NMR, since the spectral window for the pyranose protons is highly reduced (3-4.5 ppm) and 1D spectra quickly become too crowded as the number of sugar residues increases. Then, the normal routine for assigning relatively simple oligosaccharides includes the acquisition of at least 2D-^{13}C-HSQC and 2D-TOCSY spectra. The former one provides every C-H pair better dispersing the protons along the ^{13}C dimension, while the latter one permits to identify and assign each independent pyranose ring. Similar to proteins, the glycosidic linkages stop the transference of correlations, whereby the 2D-TOCSY commonly displays as many patterns as residues the oligosaccharide has.

Additionally, 2D-HMBC experiments provide ^{13}C-^{1}H correlations at longer distances (several covalent bonds) and allow to establishing the connectivity between adjacent sugars.

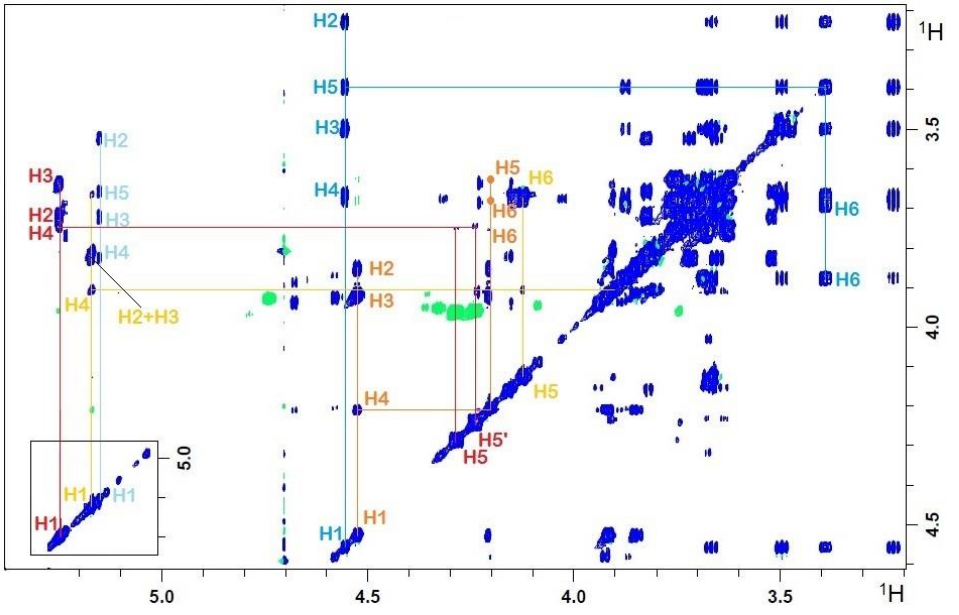


Figure 4.34. *Intrarresidue spin system correlations established via 2D-TOCSY for the B antigen. Each isolated pattern is detailed in a different color and corresponds to a specific residue. Blue: β-Glc; light blue: α-Glc; orange: β-Gal; light orange: α-Gal; red: α-Fuc.*

For the studied case, the three cited experiments (2D-^{13}C-HSQC, 2D-TOCSY and 2D-^{13}C-HMBC) were recorded at 298 K in D_2O. In both cases, the region containing the anomeric protons at low field (4.5-5.5 ppm ^{1}H; 90-100 ppm ^{13}C) showed five cross-peaks. Moreover, the chemical shift also permitted to distinguish the type of anomer (Figure 4.33): equatorial H1 protons (α anomers) are always more deshielded than axial H1 (β anomers). On this basis, three α and two β anomers were unambiguously identified. Reasonably, the additional peak was expected to arise from the Glc reducing end, whose C1 position is free in both tetrasaccharides, thereby enabling the anomeric equilibrium that gives rise to both configurations. Starting from the anomeric positions, the five correlation patterns were easily deduced from the 2D-TOCSY spectrum, in spite of the intense signal overlapping (Figure 4.34). The intensities of the TOCSY crosspeaks were primarily used to recognize the farthest positions from the reference proton (H1), thus reaching H6, which could be immediately assigned by means of their different peak signs in the ^{13}C-edited-HSQC (Figure 4.35). Only the two H6 from α-Glc turned out to be overlapped with H3 and H4 from the same residue and hence could not be

assigned in the 2D-TOCSY at first. Even so, the 2D-HMBC was more suitable for this task, especially in order to distinguish H2 and H3, which gave similar intense correlations when regarded from H1. In contrast to the 2D-TOCSY, in the 2D-HMBC the adjacent proton clearly gives the strongest correlation in most of cases. H4 was easier to correlate given that its signal in Gal rings usually appears broadened and displays very weak or no visible couplings.

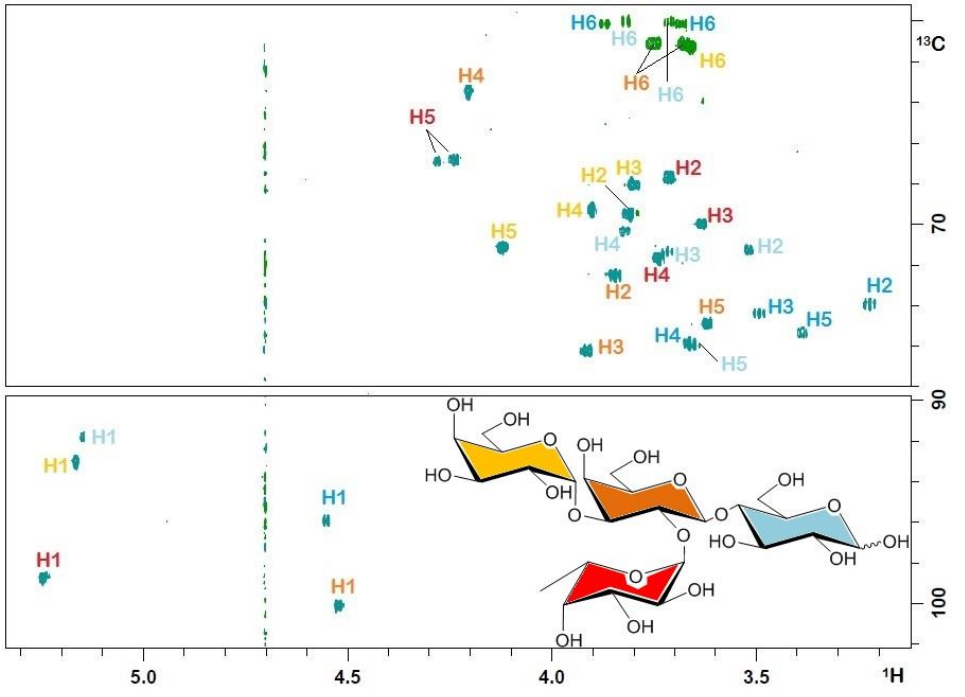


Figure 4.35. *^{13}C-HSQC full assignment of the blood group B tetrasaccharide. Blue: β-Glc; light blue: α-Glc; orange: β-Gal; light orange: α-Gal; red: α-Fuc.*

In summary, the 2D-TOCSY and the complementary 2D-HMBC were enough to assign every peak in the discerned correlation patterns. The locations of the crosspeaks along the spectral window were almost identical for both oligosaccharides, whereby both assignments were run in parallel. At this point, two particular spin systems led to eventually finish the assignment: firstly, the only set of signals that displayed correlations with the aliphatic methyl group at 1.5 ppm could be immediately recognized as the Fuc residue. To note, the anomeric equilibrium at the Glc moiety also affected to the nearby Fuc H5, giving rise to two related crosspeaks and pointing out how close this proton should be to the Glc ring. Analogously, H2 distinctively appears shifted when attached to an N-Acetyl group, as could be noticed for GlcNAc in the case of the blood group A. By discard, the remaining spin system supposed to be a Gal moiety (broad H4 singlet) was identified as the central

Gal pyranose for both antigens. Finally, the identification of the different sugars according to their glycosidic bonds was likewise checked out by means of the 2D-HMBC: each H1 could be correlated via the glycosidic oxigen with the proton attached to the opposite carbon, placed at the adjacent residue, what eventually permitted to verify the full assignment.

With this information, the saturation transfer difference experiments were carried out at 293K using the standard protein buffer prepared in D_2O (Tris-d_{11} 25 mM, NaCl 150 mM, $CaCl_2$ 4 mM, DTT 2 mM, pD=8). After several tests, the STD intensities were checked out to not significantly improve when using the EC domain of DC-SIGN, and the same experimental setup than that used for the HSQC-titrations was maintained, making use of the CRD. For a maximum ligand concentration of 4.2 mM, the CRD was added in a concentration 70 times lower (60 μM), enough to record good STD intensities while rebinding problems were conveniently suppressed. The working temperature (293K) was chosen in order to optimize the signal dispersion observed at the anomeric region, avoiding as much as possible overlapping issues with the water signal. At 298 K, the H1 protons from the β-anomers resonated too close to the water peak, but they could be efficiently pulled down to lower chemical shifts by decreasing the temperature. Conversely, the H1 of α-Gal, which is less shifted in the A antigen, approximated too much to the water signal at 288K (Figure 4.36). Noteworthy, even using commercial D_2O 99.99% pure, the remaining amount of visible water protons is predictably about 10 mM, which was within the same order than the amounts of ligand used. Of course, the only practical solution is to make use of water suppression pulses (e.g spin-lock), but no water filters were eventually included since its application was noticed to bias the STD effects of the close anomeric peaks, modifying their relative intensities.

Regarding the type of saturation, two resonances were chosen for applying the selective saturation pulse, where the CRD NMR profile contained a high percentage of signals: 7.05 ppm (aromatic) and 0.76 ppm (aliphatic). In the latter case, a good set of STD effects were reported and subsequently analyzed, whereas no clear perturbed signals could be detected when using aromatic saturation (Figure 4.37). From a structural point of view, the area surrounding the primary calcium site on DC-SIGN CRD is mainly composed of negatively charged and non-charged polar amino acids that are essential for the stabilization of the positive Ca^{2+} and the establishment of polar contacts with carbohydrates. Then, since aromatic and aliphatic residues are necessarily placed far away from the binding site (Figure 4.37), the direct

transference of saturation was in general more hindered than expected for other proteins. It seemed that aliphatic saturation was still good in such a scenario, which was attributed to the sole presence of the Leu371 and Val351 residues, close enough to the primary binding site.

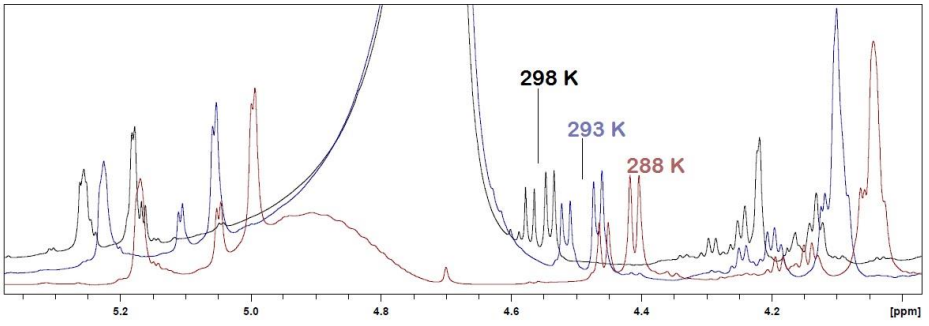


Figure 4.36. *Off-resonance experiments performed with a non-labeled CRD sample (60 μM) at three different temperatures. The spectra at 288 K and 293 K were recorded in the presence of the A antigen, while the other spectrum (298 K) shows the B antigen peaks. Water suppression was only applied in the first test (288K) to check its effect in the close ligand peaks.*

The assignment of the observed STD signals was better accomplished from a qualitative point of view. Unfortunately, the extremely intense signal overlapping observed for the studied blood groups in the recorded spectra precluded the precise observation of isolated peaks. In particular, only six protons (β-Gal H1, β-Glc H1, Fuc H1, Fuc H5, β-Glc H5 and β-Glc H2) were separated enough to be accurately integrated. Then, peak intensities were employed rather than integrals to delineate the binding epitope, as the approximation of *initial slopes* was considered unrealistic for our particular system. Accordingly, the relaxation issues could not be properly monitored as the build-up curves were neither determined in most of cases. To our knowledge, deviations from reality have been previously checked out not to be so severe for pure carbohydrates in many cases if saturation times are not excessively large [73,78-79]. The only warning in these reported cases mainly concerns the STDs of H6, which usually evolve rather differently. Then, the qualitative interpretation of the spectra was performed in those STD experiments recorded with selective saturation prolonged for 2 seconds, an intermediate delay to afford enough sensitivity without deeply biasing the relative STDs intensities. In any case, it is worth mentioning that the purpose of these experiments was exclusively descriptive, and they were meant to provide informative data that will subsequently serve for the ultimate objective, that is to assess novel theoretical binding models for describing the interaction of the studied blood antigens with DC-SIGN.

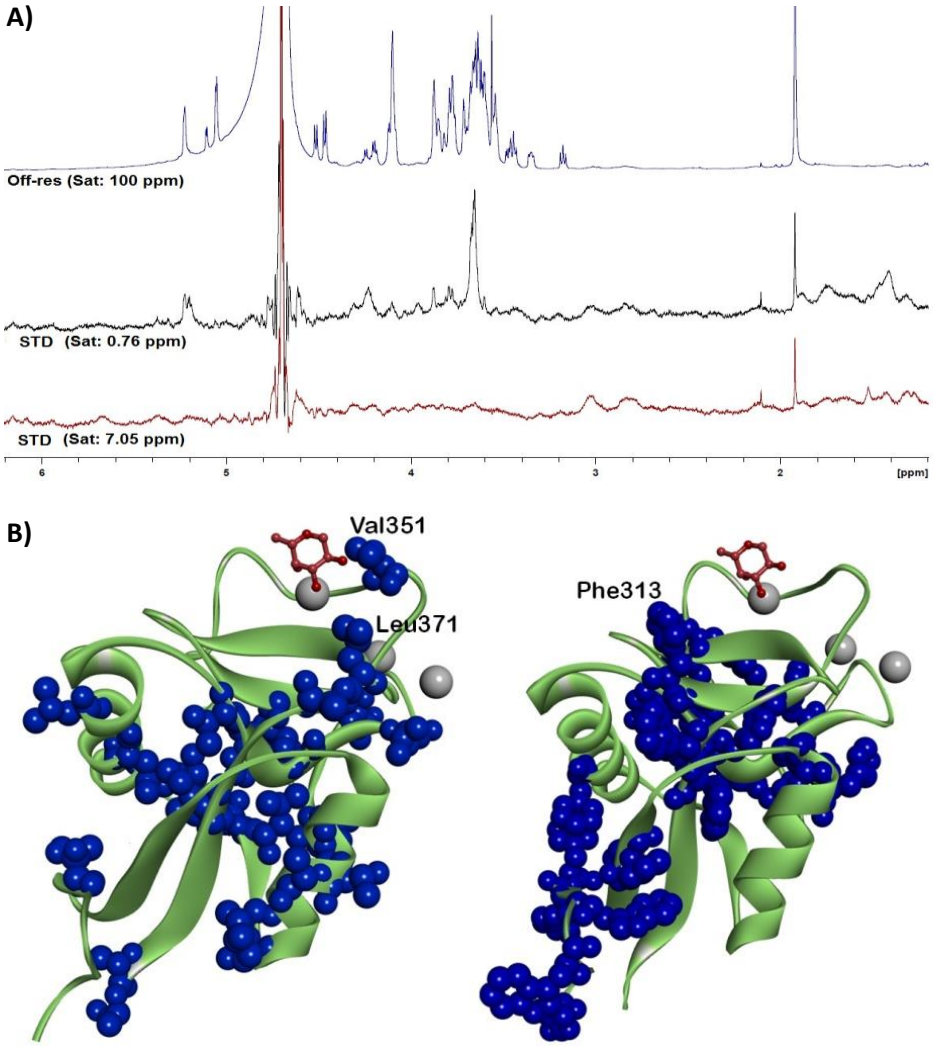


Figure 4.37. *A) STD profiles acquired at 293 K with two different saturation resonances. Top (black): aliphatic. Bottom (red): aromatic. B) Spatial distribution of aliphatic (left) and aromatic (right) residues in the DC-SIGN CRD structure. The Fuc residue from the PDB 1SL5 is represented in red.*

The global analysis of the STD spectra is detailed on Figure 4.38. Both oligosaccharides provided similar saturation profiles, sharing few characteristic STD effects that could be clearly associated to the existence of a binding epitope: the Fuc residue undoubtedly gave the strongest STD intensities, in which Fuc H2 was the most perturbed nucleus followed by Fuc H1 in both cases. This latter atom seemed to be more affected in the B antigen than in the A antigen. In absolute terms, the STDs reached 1% of intensity as much, what was rather low. Nevertheless, sensitivity issues were not a complication as acquisition was prolonged enough to improve the signal-to-

noise ratio and obtain measurable signals. Indeed, the peak overlapping was the major problem, instead of the sensitivity.

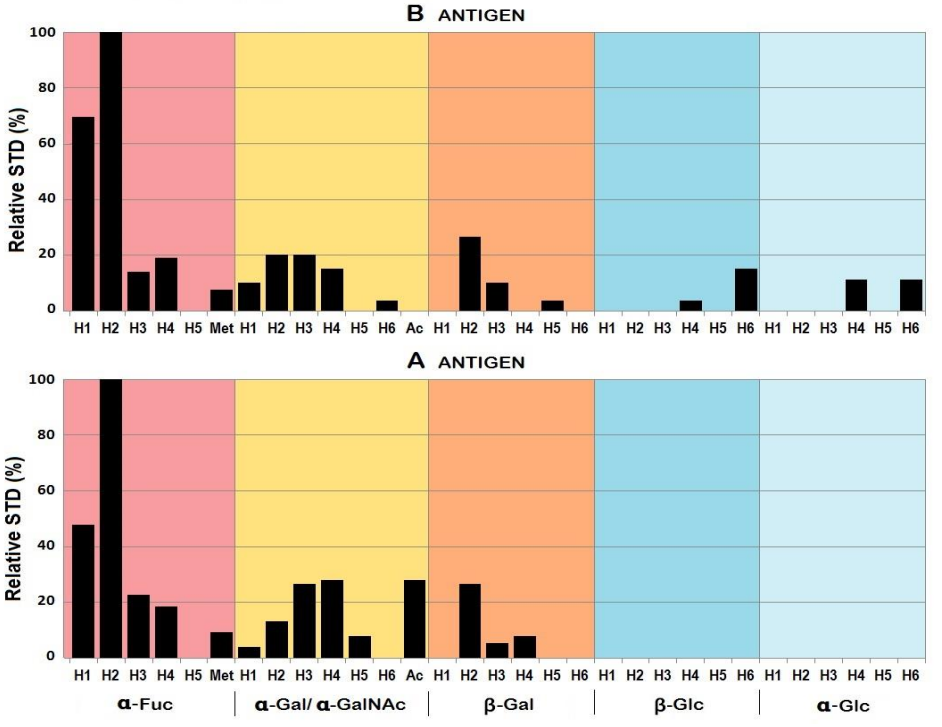


Figure 4.38. *Comparative analysis of the STD intensities obtained for both blood groups in the presence of the DC-SIGN CRD. In both cases, the Fuc H2 proton was taken as the reference proton and its relative intensity was assigned to a 100% of STD.*

Taking Fuc H2 as the reference atom (STDrel = 100%), the rest of protons were noticed to display weak relative STDs, always below 30%. H3 and H4 from Fuc also gave weak but meaningful STDs, whereas no relevant contributions were detected for H5 and the methyl group. Similarly, a significant set of notable STDs were distributed on the Gal moieties (Figure 4.39): the non-reducing Gal residue seemed to be systematically perturbed at H2, H3 and H4, and also the N-Acetyl group received an important amount of saturation in the blood group A. In addition, the central β-Gal ring showed a relatively intense STD peak for its H2 proton. Conversely, both reducing Glc anomers were almost devoid of observable STDs: for the A antigen, all the proton signals of the β-anomer were practically erased in the STD and the same was reported for H1 and H2 atoms of the α-anomer. Unfortunately, protons from H3 to H6 could not be well distinguished due to the severe signal crowding, and their exact positions could not be checked out (not measured). For the B antigen, H6 was noticed to be weakly saturated in both Glc anomers, but it is worth mentioning that

the experimental determination was very imprecise and probably overestimated in such a case, as the H6 of the β-anomer was partially overlapped with α-Gal H4 (which was more likely to yield that STD percentage), while in the α-anomer both H4 and H6 resonated at equivalent positions, and hence the STD estimation was jointly contributed by both atoms. In general, the individual STDs could be well inferred from those crowded regions since every peak displayed at least a non-overlapped region of the multiplet that could be taken to measure the peak intensity.

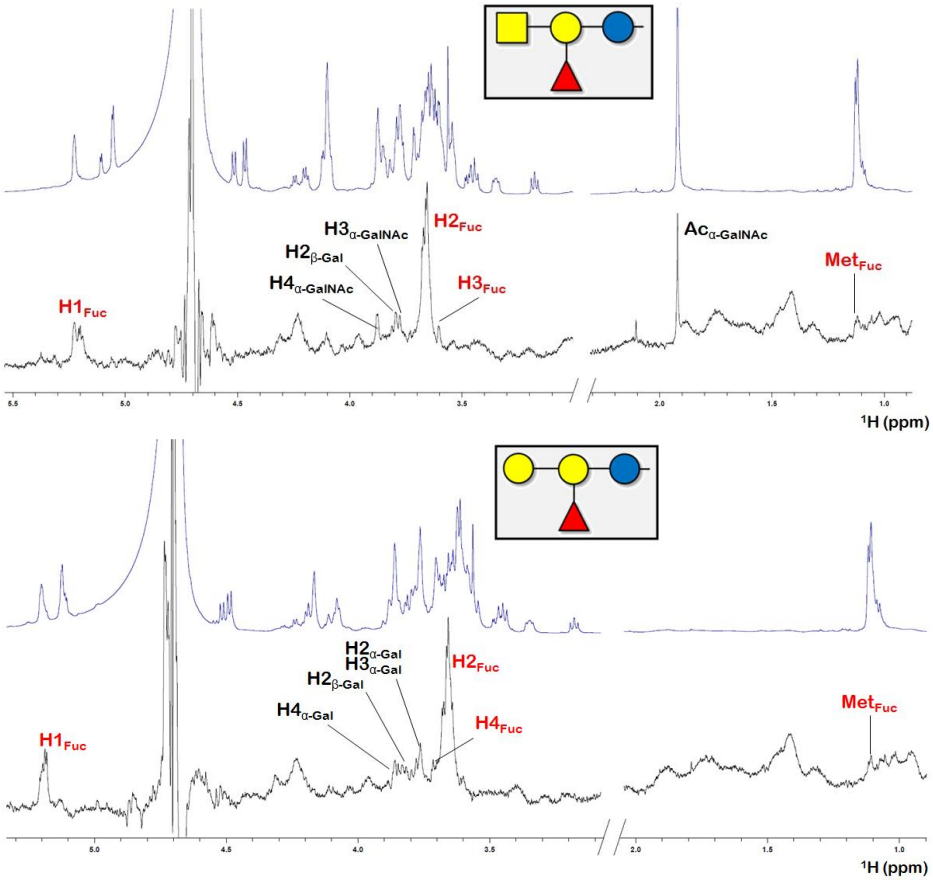


Figure 4.39. *Most relevant STD effects detected for each blood group tetrasaccharide (Top: A antigen. Bottom: B antigen). The STDs from the Fuc residue are highlighted in red.*

Overall, these STD effects probably encode the closeness of the ligands to the only aliphatic sidechains on the surroundings, namely V351 and L371. Indeed, the results strongly pointed to the almost exclusive participation of the nearby methyl groups from V351 on the saturation of the interacting antigens. The highest STD percentages measured for Fuc H2 and Fuc H1 perfectly match the

binding pose described in PDB 1SL5 for the Fuc ring. In fact, the direct vdW contact between Fuc H2 and this aliphatic sidechain has been already underlined as a key contact to form stable complexes with Lewis-type antigens [38,59,80]. Analogously, the STD values reported herein suggest that the binding pose adopted by the blood group antigens may be established in the same way than that described for the Lewis epitope, in which the calcium ion at site 2 is targeted via Fuc OH3 and OH4, such that H1 and H2 point towards Val351.

Additional information in regard to the neighboring sugars indicated that the non-reducing Gal might be rather close to the protein surface as well, while the reducing end would be presumably placed far away. As stated before, the speech related to "surface proximity" may not be strictly certain given that the Val351 is probably the main contributor to the observed STDs. Therefore, the previous reasoning could be also valid by supposing that the α-Gal moiety was close to Val351, while the reducing end could be simply contacting the protein surface at another point, far away from Val351, as reported for Man-containing oligosaccharides. However, since the configuration of the glycosidic linkages is different to that of the Lewis-type scaffolds, no straightforward conclusions could be extracted at this point on the orientation and location of the surrounding sugars. Nevertheless, the collected information turned out to be essential to further complement the data from tr-NOESY experiments, which are subsequently detailed.

To unveil the binding pose of the oligosaccharides studied in this work, tr-NOESY spectra were performed in order to derive distance restraints in the presence of the DC-SIGN CRD. At 600 MHz, the free ligand 2D-NOESY spectra displayed extremely weak positive cross peaks. This fact was actually expected since the molecular weights of the studied tetrasaccharides are within the range close to zero NOE (650 Da). At higher magnetic fields, the Larmor frequencies are larger and the relative NOE intensities may improve according to the relationship between the NOE mechanisms and the $\omega_0\tau_C$ product (See *Methods*). However, although the spectra significantly improved at 800 MHz and several negative peaks were now observable, it displayed rather few signals yet, notably attenuated in many cases. In such cases, the 2D-ROESY variant constitutes the best alternative since it always provides positive correlation peaks regardless the molecular correlation time [81]. As starting point, at 298K and 800 MHz, the ROESY experiments performed for the free ligands displayed a larger set of positive peaks that could be thoroughly examined.

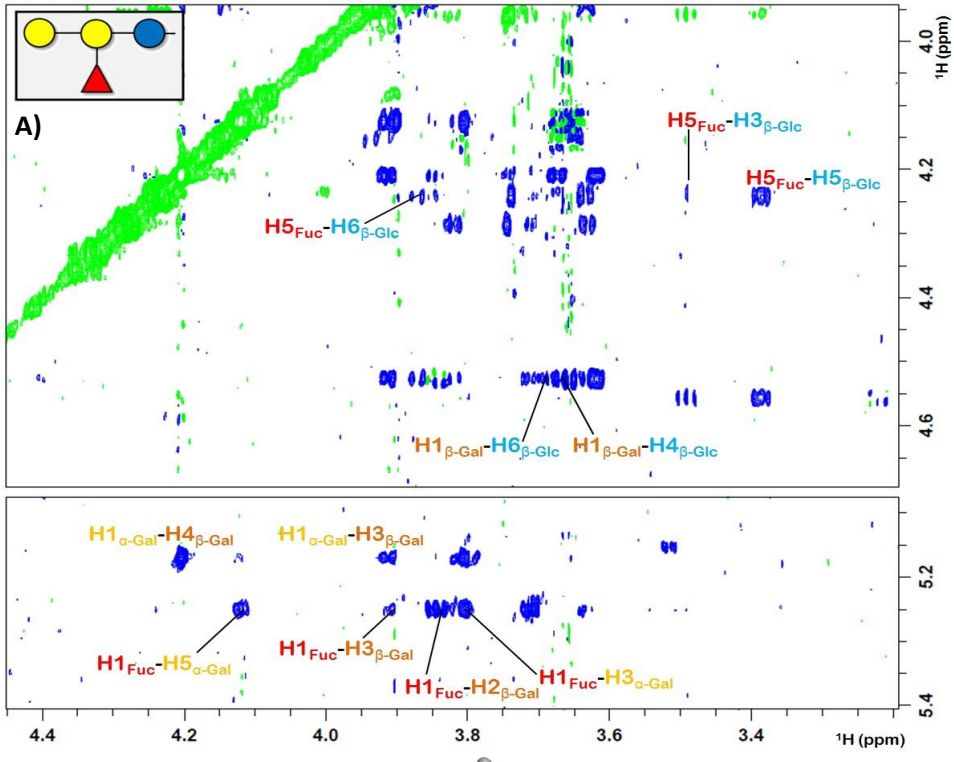
The key inter-residue contacts were carefully checked as they are essential to describe the free conformation of the tetrasaccharides (Figure 4.40A). In parallel, both ligands were modeled by means of the GLYCAM-Web online server, using the Carbohydrate builder tool. This online platform allows the prediction of the 3D structures of carbohydrates, making use of the "GLYCAM06" force field (latest update) developed and optimized by the team of Prof. R. J. Woods [82-83].

The NOE distances were estimated and compared to those measured in the modeled structures. Experimentally, there are several methods to establish the relationship between NOE intensities and inter-protonic distances [84], being the full relaxation matrix approach one of the most rigorous ones. However, this method is usually employed when studying large molecules in which spin-diffusion effects can easily lead to erroneous estimations. Instead, the approaches based on relative distances are approximate but still precise enough to provide interatomic distances in small relatively rigid molecules [85-86]. Using a known distance, the unknown one is deduced from the expression on Eq. 2, which correlates the quotient between both distances with their respective cross relaxation rates (and therefore, indirectly with the NOE peak integrals):

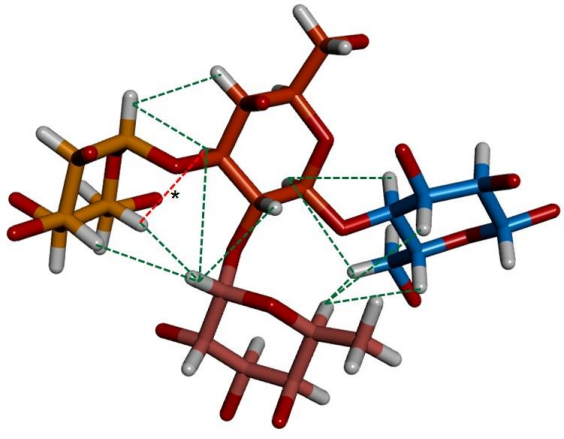
$$\frac{\mathrm{NOE_{ij}}}{\mathrm{NOE_{ab}}} \cong \frac{\sigma_{ij}}{\sigma_{ab}} = \frac{\mathbf{r_{ij}}^{-6}}{\mathbf{r_{ab}}^{-6}} \qquad (Eq.2)$$

Where "a" and "b" are the two spins whose interatomic distance is known a priori, and "i" and "j" form the other spin pair which is being evaluated. Thus, the collected interresidual NOEs were all processed by using Eq. 2 and three intrarresidual NOE effects: H1-H3 β-Gal, H1-H5 β-Gal and H3-H5 β-Glc. In β-Glc and β-Gal residues, these protons are all placed at the bottom plane of the sugar structure, and their relative distances have been taken as fixed values (ca. 2.4 Å), given the exclusive $^4C_1$ conformation adopted by these D-sugars in solution [87]. The resulting average interatomic distances are summarized in column 2 of Figure 4.41. All values were slightly underestimated compared to those predicted by the GLYCAM model. However, the relative errors were mostly ranging around 8-14%, what could be attributed to a systematic error, probably arising from the experimental determination of the peak integrals with the NMR software. Most importantly, from a qualitative perspective, all the interresidual NOEs found fitted to those expected for the modeled tetrasaccharides (Figure 4.40B), except for the αGal H5 - βGal H3 cross peak, whose NOE correlation could not be detected in any case despite the distance

in the GLYCAM model was quite short (2.5 Å). Slight local angular motions of the corresponding lingakes could be responsible for this fact. In summary, the theoretical model was eventually taken as valid, after verifying that the glycosidic torsion angles were within the expected values for the exo-anomeric effects.



B)



	A antigen		B antigen	
	φ	ψ	φ	ψ
α-Gal/GalNAc	-51.4	-35.4	-69.0	-36.6
β-Gal	+53.8	+5.9	+85.7	-27.2
α-Fuc	+49.8	-4.4	+68.9	+12.7

Figure 4.40. *A) 2D-ROESY spectrum for the blood group B antigen (1 mM/D_2O, t_{mix}=200 ms), showing the assigned interresidual NOEs. B) The free conformation of the B antigen, as predicted by GLYCAM06. The experimental NOEs are depicted as dashed green lines. The red line with an asterisk points out the short distance H5 α-Gal-H3 β-Gal, not found in the 2D-ROESY. The table below includes the corresponding dihedral angles φ and ψ observed for the three torsional glycosidic linkages in each tetrasaccharide.*

Subsequently, tr-NOESY spectra were recorded with both ligands in the presence of the lectin. Both experiments were acquired with a 1:5 lectin:ligand ratio and 200 ms mixing time, enough to accumulate NOE without significant spin diffusion. At 298 K, a remarkable increase in the negative NOE intensities was noticed for both ligands, thus confirming their low molecular tumbling due to specific binding to the lectin.

	GLYCAM model	2D-ROESY	ε_R	Tr-NOESY (1:5 CRD/BgB)
$H1_{Fuc}$-$H3_{\alpha Gal}$	2.5	2.3-2.4	8 %	-
$H1_{Fuc}$-$H5_{\alpha Gal}$	2.8	2.5-2.6	10%	2.8-3.2
$H1_{Fuc}$-$H2_{\beta Gal}$	2.7	2.3	14%	-
$H1_{Fuc}$-$H3_{\beta Gal}$	3.9	2.9-3.0	25%*	2.6-3.0
$H1_{\alpha Gal}$-$H3_{\beta Gal}$	2.9	2.6-2.7	10%	2.5-2.8
$H1_{\alpha Gal}$-$H4_{\beta Gal}$	2.3	2.0-2.1	13%	2.1-2.4
$H5_{\alpha Gal}$-$H3_{\beta Gal}$	2.5	-	-	-
$H1_{\beta Gal}$-$H4_{\beta Glc}$	2.7	2.3	14%	-
$H1_{\beta Gal}$-$H6_{\beta Glc}$	3.3	2.5-2.6	24%*	-
$H5_{Fuc}$-$H3_{\beta Glc}$	3.3	2.5-2.6	24%*	-
$H5_{Fuc}$-$H5_{\beta Glc}$	2.9	2.7-2.8	7%	2.8-3.2
$H5_{Fuc}$-$H6_{\beta Glc}$	3.4	3.2	8%	-

Figure 4.41. *Proton-proton distances (Å) derived from computational calculations (GLYCAM) and experimental data in free solution (2D-ROESY) and in the presence of 0.2 eq of DC-SIGN CRD (Tr-NOESY). The relative errors (ε_R) are calculated respect to the values in the first two columns. The asterisks highlight the largest distances, for which the errors were higher.*

The strong negative NOEs were thoroughly scrutinized to search for substantial differences between the free and the bound antigens. Interestingly, the set of transferred NOEs was nearly equivalent to that obtained for the free ligands (Figure 4.41, *last column*). At least five peak integrals could be calculated and their values also resulted in distances (See *Methods*) within the same range as those of the free ligands, except for Fuc H1 - αGal H5, whose inter-proton distance significantly increased. To note, the

peak intensities improved but they turned substantially broader due to the free-bound exchange process, increasing the peak overlapping and precluding the proper quantification of more cross peaks. In any case, the obtained data strongly suggested that both ligands essentially maintained their respective free conformations in the complex with the DC-SIGN CRD, with minor variations in the inter-proton distances, and especially considering a slight torsional twist on the Gal-Gal glycosidic linkage to account for the change in the Fuc H1 - αGal H5 distance without altering the other ones.

Apart from the information related to the inner correlations in the ligand structure, a more detailed inspection of the tr-NOESY spectrum revealed an interesting feature. H1 Fuc was noticed to present a new correlation with some nuclei at very high field, around 0.80 ppm. Such a correlation did not appear either in the 2D-NOESY of the free protein or in the 2D-ROESY of the free ligand, thereby corresponding to an intermolecular NOE effect arising from a close contact between H1 Fuc and some protein sidechain. Additionally, the tr-NOESY yielded two more correlations with the same unknown nucleus arising from H2 Fuc and, in a lesser extent, from H3 α-Gal (Figure 4.42). From the protein perspective, the only sidechains expected to have high shielded protons are those belonging to the aliphatic residues, namely Leu, Val and Ile [88]. Fittingly, the unique residue included in this group that was located close to the ligand in the CRD architecture was actually Val351. Indeed, the measured chemical shift fitted very well with those reported for methyl groups from valine residues [89-90] (Figure 3.7 on *Chapter 3*).

Obviously, the veracity of the aforementioned conclusion could not be checked out exclusively through 2D-NOESY experiments with such a severe signal crowding. To prove the existence of these close contacts, a 3D-^{15}N-HSQC-TOCSY was performed with a sample containing the ^{15}N-labeled CRD in the same experimental conditions, namely, 298 K and 5 eq of ligand. The experiment is equivalent to a 2D-TOCSY spectrum in which the individual TOCSY patterns are further separated in the additional nitrogen dimension, thereby notably reducing the signal overlapping. When positioning at 121.4 ppm in the ^{15}N dimension, the TOCSY plane showed a correlation pattern for a peak located at 7.15 ppm, which corresponded to the Val351 amide group (Figure 4.43). The TOCSY crosspeaks, from high to low field, were respectively identified as correlations with protons Hγ (methyl groups), Hβ and Hα. Not only the Hγ chemical shifts were in perfect agreement with those identified in the tr-NOESY, but also the other correlations could be unequivocally identified

as intra-residue NOEs. Overall, although the signal crowding precluded the quantitative estimation of intermolecular distances, this piece of information provided strong spatial constrains to restrict the quest for a suitable binding pose, given that NOEs cannot be generally observed above distances around 5 Å.

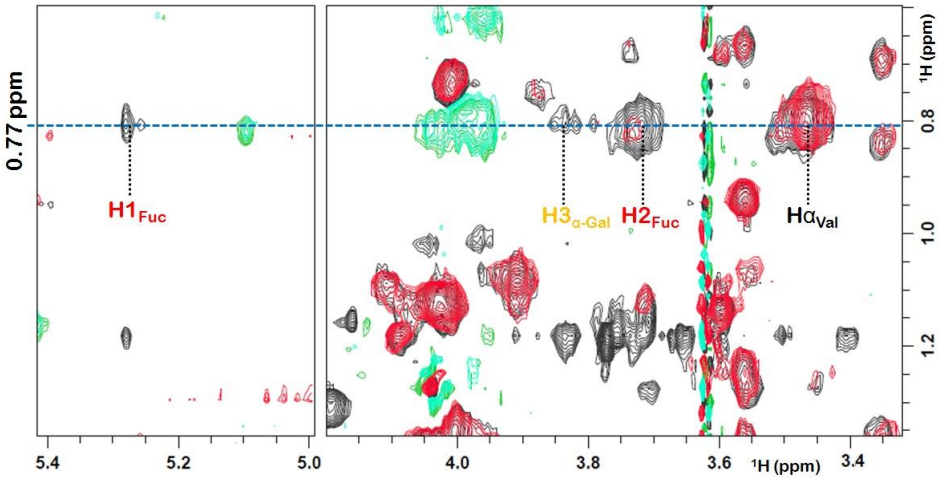


Figure 4.42. *In black, the tr-NOESY of the blood group A in the presence of 0.2 eq of lectin CRD (t_{mix} = 200 ms). In red, a 2D-NOESY of the free protein performed with the same experimental setup (t_{mix} = 200 ms). As highlighted, the correlations not observed in the free NOESY involve contacts with the ligand protons.*

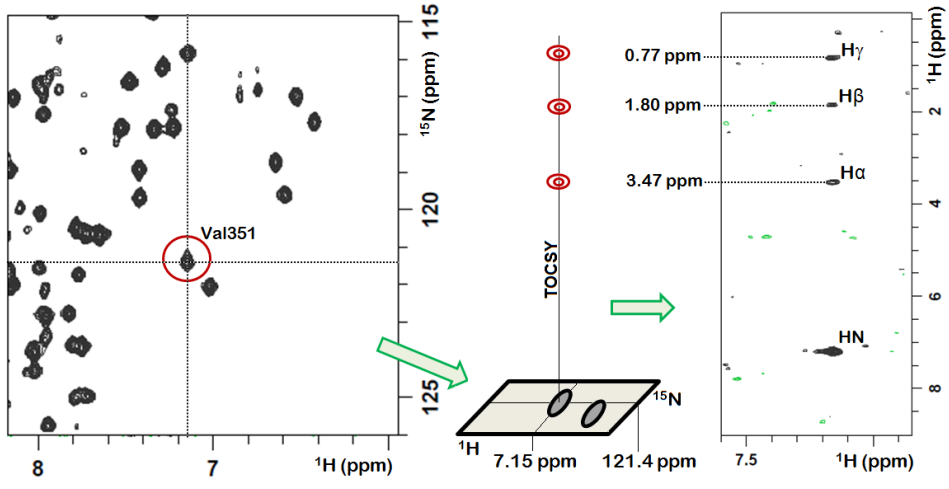


Figure 4.43. *Sidechain assignment of the residue Val351 through 3D-^{15}N-HSQC-TOCSY at 298 K, in the presence of 5 eq of A antigen (t_{mix} = 80 ms). On the left, the ^{15}N-HSQC projection, and on the right, the TOCSY plane extracted at δ(^{15}N) = 121.4 ppm, showing a strip centered on 7.15 ppm in the ^{1}H dimension.*

At this point, the gathered information was enough to propose a first 3D model of the sugar/lectin complex. From the STD results, it was concluded that the Fuc residue was likely to be oriented such that H1 and H2 directly contact the methyl group of Val351. Following this reasoning, the Fuc moiety was supposed to interact at the canonical binding site through OH3 and OH4, as it does in the PDB 1SL5, thereby fulfilling these experimental findings. Moreover, as deduced from the tr-NOESY data, the spatial arrangement of the other sugar residues adopted in the structure of the bound histo blood group antigens corresponded to that described by the theoretical models presented above (Figure 4.40B). Then, the 1SL5 coordinates were employed to fit the structures of the bound tetrasaccharides by means of superimposing the Fuc ring of both antigens with the Fuc ring of the LNFP III oligosaccharide. The resulting binding poses are detailed in Figure 4.44.

In general terms, for both antigens, the proposed model nicely fulfilled the ligand-derived information described above. The largest STD effects were almost perfectly distributed among those ligand protons surrounding the Val351 sidechain. According to the crystal model, the methyl groups of this residue are very close to Fuc H1 (3.4 Å) and Fuc H2 (3.0 Å). Fittingly, the α-Gal moiety was placed fairly close to Val351 as well, thereby supporting the observed saturation of this residue mainly contributed by H2, H3, and H4. Analogously, for the core β-Gal, H2 was the closest proton to this amino acid, what would justify the only relevant STD effect observed. Finally, the proposed arrangement would place the remaining Glc away from both the protein surface and specifically from Val351, supporting the almost complete absence of STDs. In regard to the tr-NOESY data, the interatomic distances between Hγ of Val351 and Fuc H1, Fuc H2 and α-Gal H3 take values below 5 Å in the models, thus fitting with the experimental NOEs reported. Such a finding constituted the strongest proof to support the modeled binding pose, as the insertion of the Fuc ring into the binding site through other contacts (OH2/OH3, as in Langerin [74], 180º-rotated OH3/OH4, as in DC-SIGN interacting with Man [60,73]), would inevitably place at least one of the three mentioned protons too far from Val351.

Finally, it is worth mentioning that this approximation constituted a preliminary model roughly built by manual docking, which was subsequently refined by molecular dynamics (MD) simulations, as described in the following section.

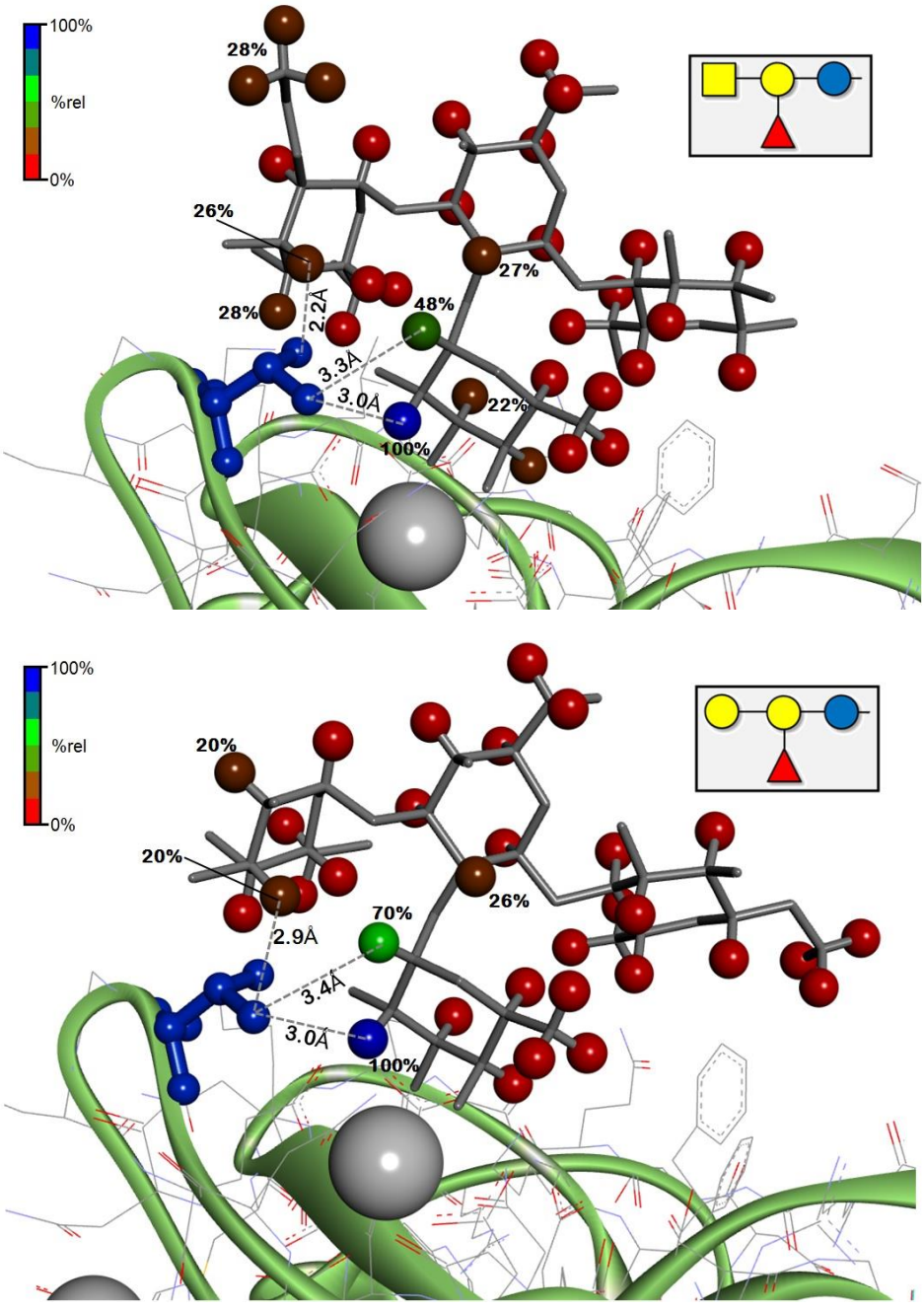


Figure 4.44. *Binding poses proposed for the blood groups A (top) and B (bottom) on the basis of the collected STD and tr-NOESY data. Val351 is represented in blue. The protons are colored according to their respective relative STD percentages, from blue (100%) to red (0%). Relative STDs higher than 20% are directly indicated. The three transferred NOEs between Val351 methyl groups and the ligand are detailed as well (the protons of the Val351 residue are not represented to simplify the figure).*

4.5. Molecular Dynamics

Once a suitable description of the interaction had been proposed, the theoretical binding models were subjected to molecular dynamics simulations, aimed at assessing their stability and getting additional insights into the interactions. For both oligosaccharides, the starting geometry was built as detailed above, namely, by superimposing the Fuc ring from the ligands with the same residue from the LNFP III scaffold present in the PDB 1SL5 structure. After the two initial minimization steps, the MD simulations were run for 100 ns. Along this time, structural data were extracted every 2 ps, collecting a total of 50000 frames that were carefully analyzed.

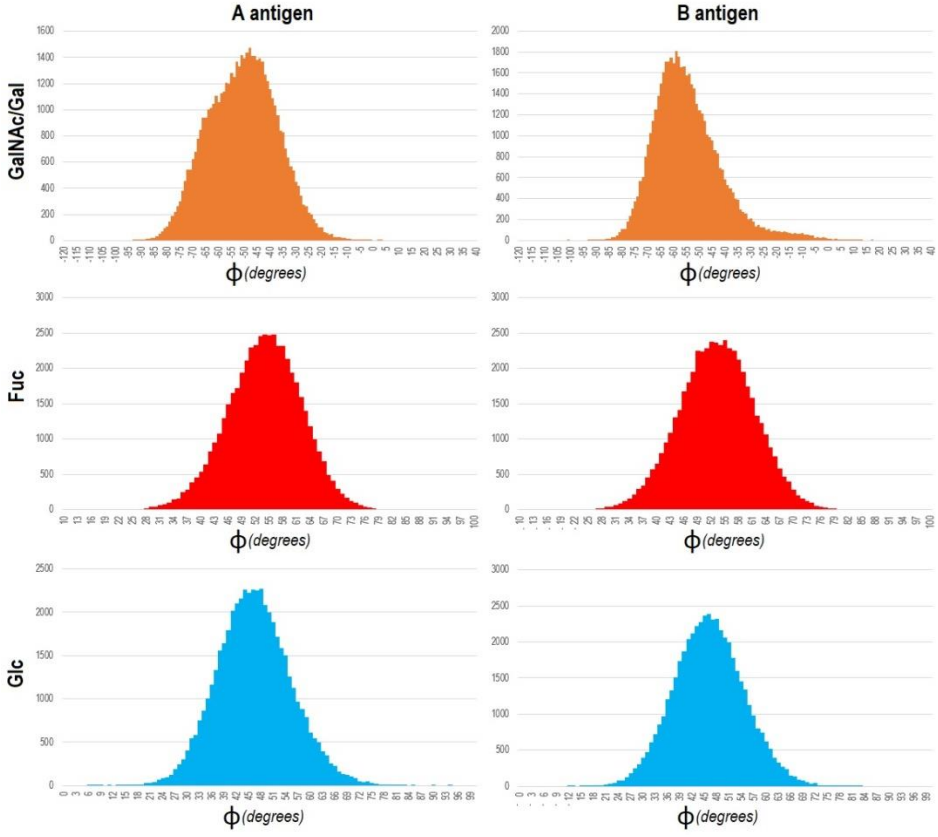


Figure 4.45. *Values of the torsional angle φ observed for the three glycosidic linkages existing in each tetrasaccharide along the complete MD simulation (100 ns). In all cases, the φ dihedral is defined from the anomeric H1 to the opposite carbon (H1-C1-O'-C'). The vertical axis represent the frequency of each dihedral value, in number of MD frames (total = 50000 points).*

Fittingly, both complexes remained stable along the entire simulation, and no ligand detaching was observed at any moment. The torsional angles ϕ and ψ

were monitored for the three glycosidic linkages. In both cases, the observed values of the ϕ torsional reasonably accounted for the expected exo-anomeric effects. Specifically, the average values corresponded in the three cases to an *exo-syn* conformation, ranging around ϕ = -53º for the α-Gal/GalNAc, ϕ = +46º for the core β-Gal and ϕ = +53º for the α-Fuc (Figure 4.45). To clarify, the sign of the ϕ angle for the *exo-syn* α-Fuc is the opposite as it is an L-sugar. Noteworthy, as can be appreciated in Figures 4.46 and 4.47, the torsional angles related to the Fucα1-2Gal linkage were very low dispersed, with standard deviations below 10 degrees. Moreover, the Fuc moiety remained almost at the same position at every moment, showing a much restricted mobility as a result of the two coordination bonds established with the calcium ion. Both observations highlighted the marked rigidity of these two residues (α-Fuc and β-Gal), barely displaced from the initial positions along the simulation. For the reducing Glc ring, the ϕ angle also fluctuated within a reduced range, while the standard deviation for the other dihedral ψ was higher, thereby conferring to this residue a broader degree of mobility.

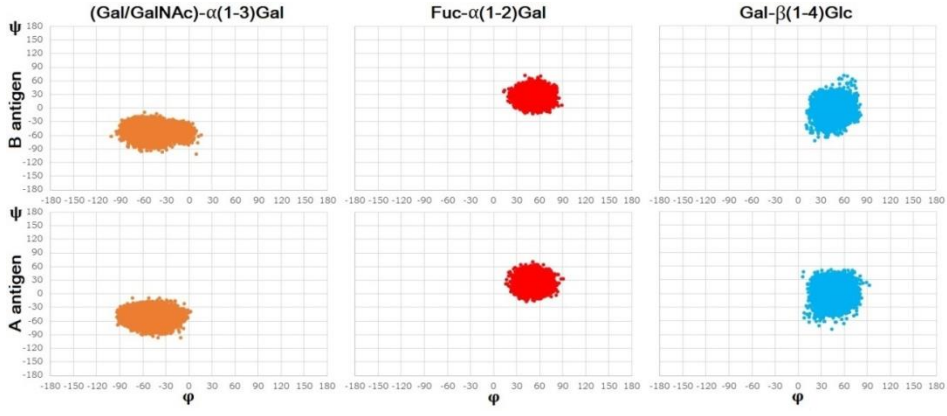


Figure 4.46. *Distribution of the φ and ψ torsional angles which define each glycosidic bond in the studied antigens. The φ dihedral is defined from the anomeric H1 to the opposite carbon (H1-C1-O'-C') and the ψ dihedral, from the anomeric C1 to the opposite proton (C1-O'-C'-H').*

Regarding the other terminal non-reducing moiety (Gal/GalNAc), the overall dynamic behavior was similar although some slight differences were noticed between the two tetrasaccharides. In both cases, the values of the ϕ angle were more dispersed, ranging from -25º to -75º for more than 95% of the collected frames. This variability led the Gal/GalNAc pyranose to show a more undefined location along the simulations (Figure 4.47, *bottom*), which could be explained by the existence of two preferred conformations of the Gal/GalNAc ring, as deduced from the asymmetric distribution of the ϕ values.

Indeed, such an asymmetry was better observed for the A antigen, which did not show a simple Gaussian distribution (Figure 4.45). In one of the conformations, the α-Gal ring displayed a ϕ value centered around -60º, which places OH3 and OH4 polar groups closer to the protein surface. In such an arrangement, the α-Gal residue forms hydrogen bonds with the carbonyl group of Val351 and the amino group of Lys368, mainly through OH3 and occasionally through OH4 as well (Figure 4.48A). In the alternative conformation, the ϕ dihedral switched to a smaller value of -47º in average, similar to the torsional angle found in the free oligosaccharides. As a result, OH3 and OH4 were slightly moved away from the previous positions, getting closer to the long loop. However, in this case, only the OH6 group of α-Gal was directly involved in polar contacts with the protein sidechains, participating in a wider network of hydrogen bonds. These contacts mediated by OH6 were alternatively established with the carbonyl group of Val351 and the charged sidechain groups of Lys368 and Glu354. Frequently, an additional intramolecular H-bond with OH2 of Fuc also bridged this residue to the mentioned amino acids (Figure 4.48A).

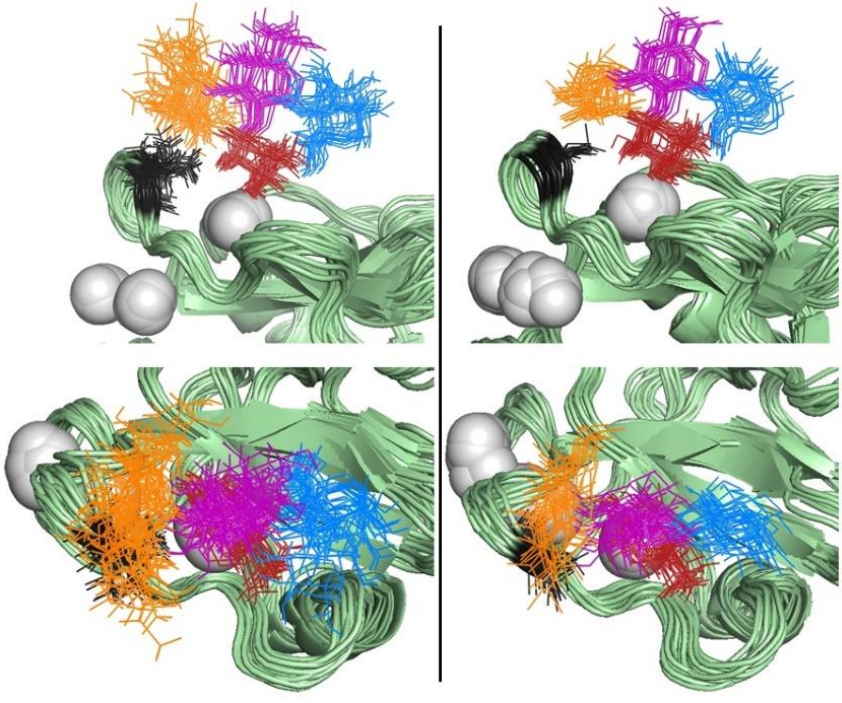


Figure 4.47. *Front and top views of the ligand trajectories (left: A antigen; right: B antigen) along the MD simulations. The bundle is constituted by 20 representative frames. Color code: blue: β-Glc; purple: β-Gal; red: α-Fuc; orange: α-Gal/GalNAc. The white spheres represent the calcium ions. The residue Val351 is detailed in black.*

Remarkably, the α-Gal/GalNAc moiety was the only residue as well as the α-Fuc moiety that established stabilizing contacts with the protein, whereas the core β-Gal and the reducing Glc were completely devoid of direct intermolecular interactions. As indicated above, the non-reducing Gal/GalNAc was involved in transitory VdW contacts with the three aforementioned residues (Val351, Lys368, Glu354), while the Fuc pyranose, as well as coordinating the calcium ion, formed fixed hydrogen bonds with at least five amino acids located on the long loop and the opposite β_4 strand: Glu347, Asn349, Glu354, Asn365 and Asp367. All these residues simultaneously took part in the coordination sphere of the metallic atom and interacted with the hydroxyl groups OH2, OH3 and OH4 from Fuc (Figure 4.48B). Worth noting, H1 Fuc and H2 Fuc remained notably close to the Val351 sidechain at every moment (2.0-3.0 Å), perfectly compiling with the intermolecular NOE effects previously reported for them. Similarly, the tr-NOE arising from H3 α-Gal could be explained as the distance to the Val351 methyl groups was always lower than 5 Å regardless the linkage conformation (Figure 4.49). The average inter-protonic distances were 3.8-4.0 Å from H1 Fuc and H3 α-Gal, while H2 Fuc was undoubtedly placed nearer Val351 (3.0-3.2 Å). These measurements would explain why the average tr-NOE effects were much weaker for the former protons. Apart, H4 also lied rather close to Val351 (2.5-3.5 Å) in one conformation, but it was completely pushed away from the aliphatic sidechain in the other conformer, reaching distances much larger than 5 Å. Thus, the averaged transferred NOE, if it exists, would be expected to be even weaker than that of H3 Gal/GalNAc, and in fact, it was not observed for any antigen at all. Taken together, these observations fulfilled with the experimental NMR data.

In addition, the corresponding MD trajectories were analyzed by using the CORCEMA-ST software, with the purpose ofevaluating the agreement between the theoretical STDs and the experimental ones. To reduce the time costs, a series of previous tests allowed to checking out that 100 structures randomly selected were enough to represent the binding of both antigens along the total MD time, since variations in the final theoretical STDs were actually negligible when more frames were included in the analysis. For each proton, its theoretical STD was calculated as the average of the 100 STDs estimated from the 100 respective computed frames. Thus, the final STD profile was built on the basis of the averaged values individually calculated and compared with the experimental profile (Figure 4.50).

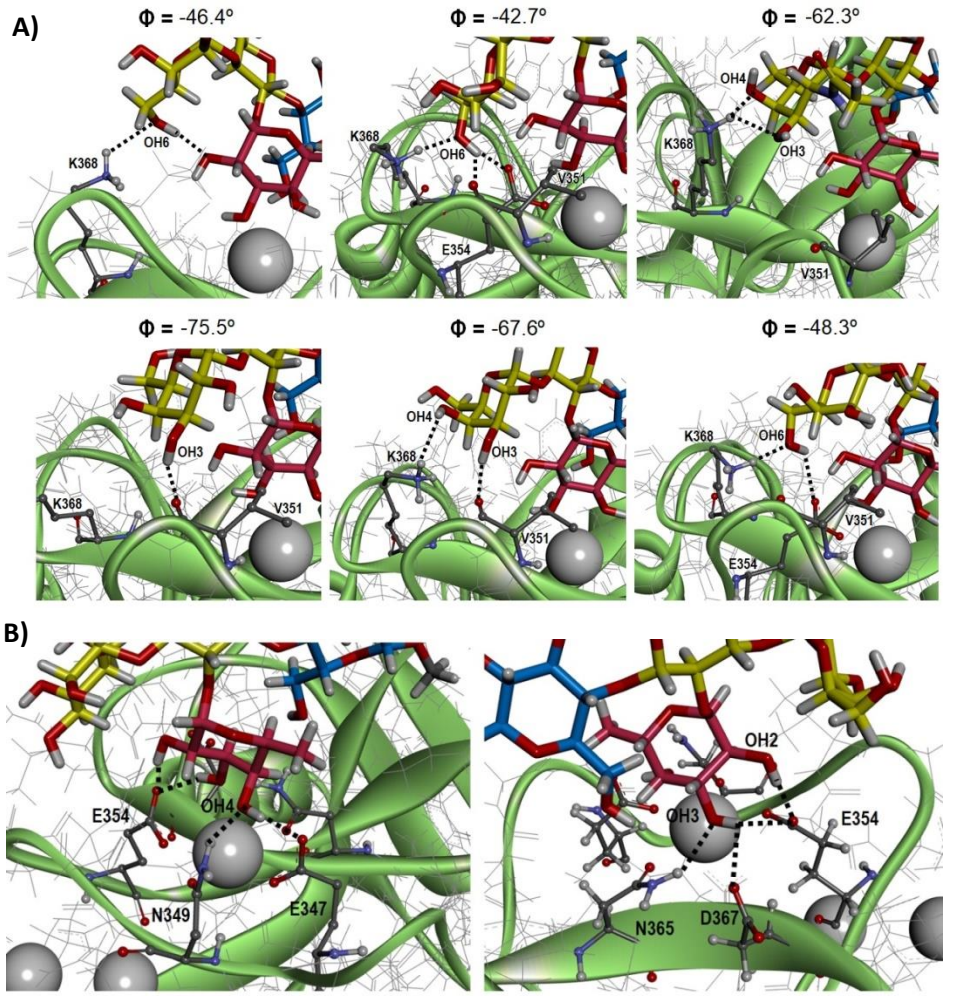


Figure 4.48. *A) Detailed analysis of the molecular interactions (H bonds, dashed black lines) in which the non-reducing Gal/GalNAc moiety is directly involved. For each antigen, three representative frames are shown (top row: BgA; bottom row: BgB) and the corresponding φ angle is indicated on the top. Also, the three protein residues participating in these contacts are highlighted as tighter lines. B) Front and back views of the binding site displaying the molecular interactions provided by the Fuc residue. The coordination bonds are not represented.*

Several experimental variables were directly introduced as known parameters for running CORCEMA-ST, whereas other conditions could not be described exactly like in the experiments and had to be estimated considering some assumptions. For instance, as the sidechain assignment was not available for the DC-SIGN CRD, no chemical shifts could be used to simulate the selective saturation pulse, and instead, this pulse was considered as entirely applied to five aliphatic amino acids (Ile, Leu, Val, and also Ala and Thr). Similarly, the

correlation time of the bound ligand (τ_{bound}) was approximately determined through the empirical relationship described by Eq. 3 [91]. Finally, the chosen values of k_{on} and K_d were those previously obtained from ^{15}N-HSQC-based titrations. To note, the on rate turned out to be rather important in the calculations, as the results significantly deviated when supposing a diffusion controlled k_{on} of 10^8 M^{-1} s^{-1} (not shown).

$$\tau_C(\mathrm{ns}) \approx \frac{\rho \cdot \mathrm{MW}}{2400} \qquad (Eq.3)$$

Where "ρ" is given the value 1 for proteins with spherical (globular) shape, or alternatively 1.32 for other irregular geometries (ellipsoidal, lineal…). The molecular weight in this case corresponds to the sum of the respective protein and ligand MWs.

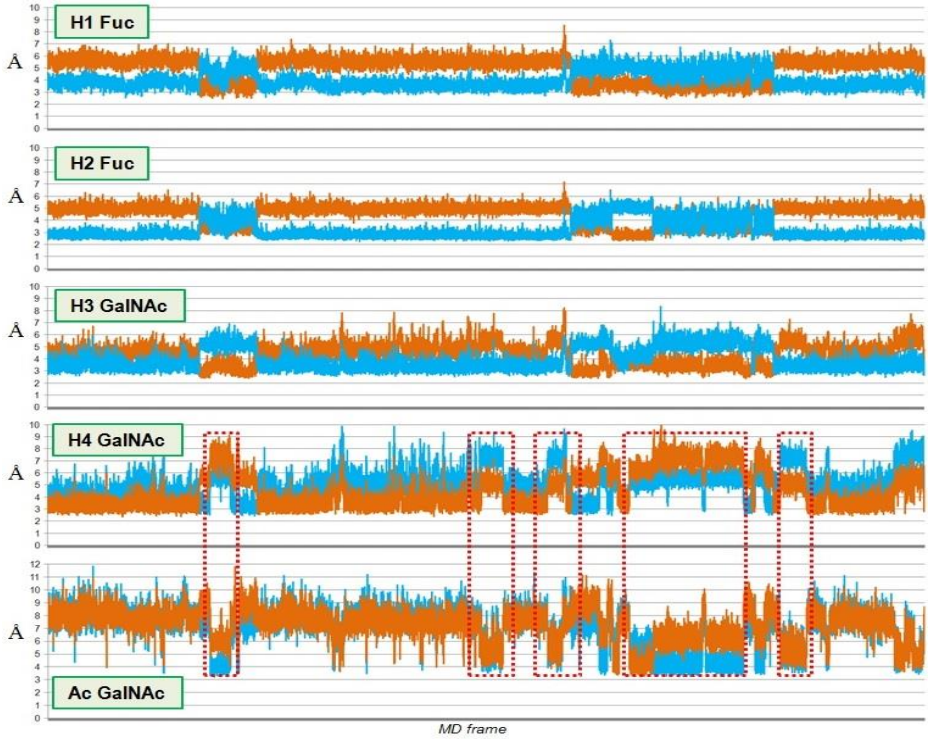


Figure 4.49. *Detailed analysis of the intermolecular distances between the two methyl groups of Val351 (blue/orange) and five protons of the blood group A. As can be seen, the distances to H1 Fuc, H2 Fuc and H3 GalNAc remain at fixed values below 4 Angstroms, always contacting at least one methyl group. For H4 GalNAc and the N-acetyl protons, the lowest distance to the Val351 sidechain systematically changes several times along the MD simulation, as a result of an important switch in the GalNAc arrangement. The red boxes indicate those time intervals in which the initial pose is modified (φ increases).*

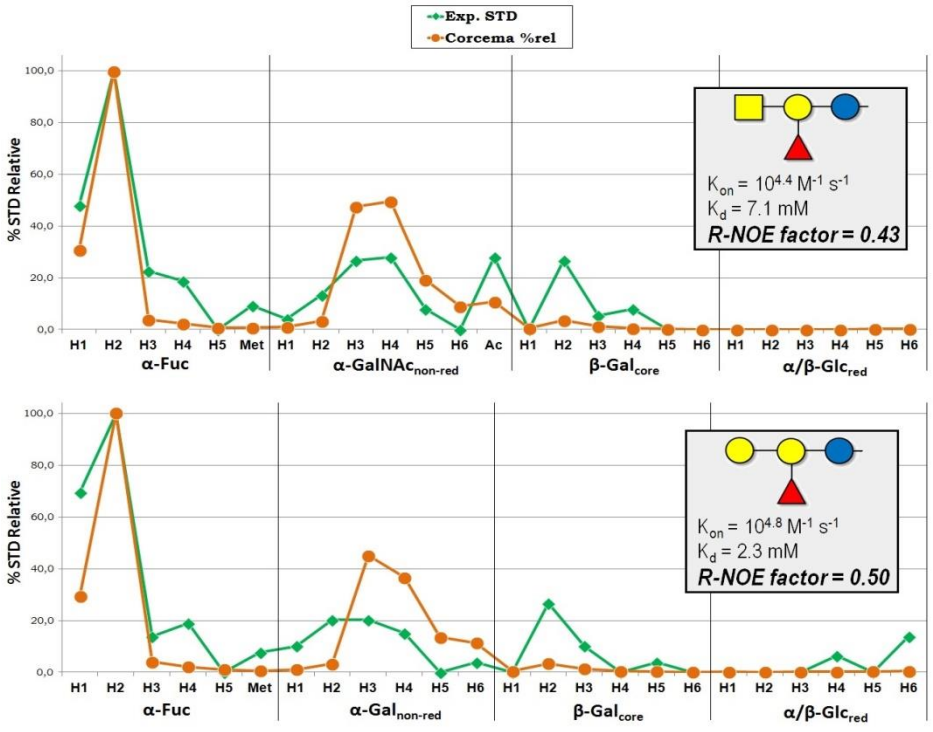


Figure 4.50. *Comparison between the CORCEMA-ST-derived STDs (orange) and the experimental STDs (green), expressed as relative values (Ref. H2 Fuc in both cases). Both CORCEMA-ST calculations were run with the following common parameters: Δ_{sat}: Ile, Leu, Val, Ala, Thr; t_{sat} = 2 sec.; $[P_0]$ = 60 μM; $[L_0]$ = 4.2 mM; rholeak = 0.1; B_0 = 600 MHz; τ_{free} = 0.5 ns; τ_{bound} = 8.33 ns; τ_{intra} = 0.01 ns.*

The resulting CORCEMA-ST profile described the experimental STDs in a satisfactory manner, although some discrepancies were found for both antigens. To highlight, the successive runs remarkably underestimated the STDs of H1 α-Fuc and H2 β-Gal, as well as the STD effect provided by the N-acetyl group in the case of the A antigen. Even so, the main differences between both datasets were observed for the α-Gal/GalNAc moiety, whose theoretical values were rather overestimated for most of their protons in comparison to those experimentally collected. However, a thorough analysis of the computed frames revealed that such an overestimation essentially came from the contribution of those frames in which the α-Gal/GalNAc moiety adopted conformations with ϕ values smaller than -60º, thus placing H3 and H4 directly towards the Val351 methyl groups. In fact, when considering any of the other frames with the ϕ dihedral close to -60º or higher, the calculated profile improved for α-Gal/GalNAc, getting closer to the experimental STD effects (Figure 4.51): the contributions from H3 α-Gal and H4 α-Gal turned weaker, while the more flexible N-acetyl group (for the A antigen) approached

to the protein surface, specifically to the Val351 sidechain region, providing a stronger average theoretical STD. In either case, the R-NOE factors (Figure 4.50, see *Methods*) were acceptable considering that no experimental restrictions were imposed during the simulations and no further optimizations were applied to the CORCEMA-ST setup, except for k_{on} [92-93].

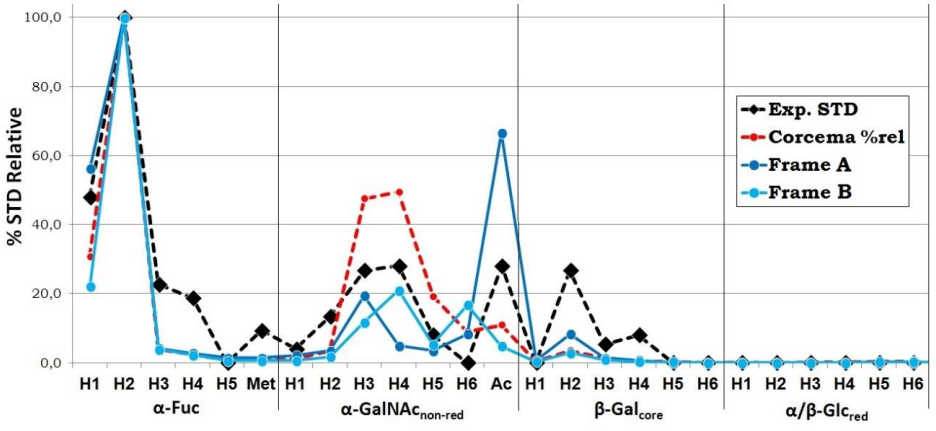


Figure 4.51. *Comparison between the average theoretical STDs obtained for the A antigen (red dashed line) and CORCEMA-ST profiles calculated for two isolated frames, A and B, in which the GalNAc moiety displayed φ angles of -65.8º and -62.3º, respectively (blue solid lines). The black dashed line corresponds to the experimental profile. As can be noted, the predicted STDs for the two selected frames are notably lower for H3 and H4 GalNAc in comparison with the average values. In contrast, the expected STD for the N-acetyl group significantly varies among frames with similar poses.*

Therefore, the NMR results presented herein suggested that in the real scenario the α-Gal/GalNAc pyranose is likely to show a fixed conformation, exhibiting a ϕ angle close or higher than -60º, but not lower. The alternative conformer, in the most extreme cases, led this pyranose ring to pack against the nearby α-Fuc residue, thereby enabling hydrophobic contacts between both sugars as well as with the sandwiched aliphatic residue (Val351), what was not fully supported by the experimental NMR data. Hence, such an event was thought not to be actually happening, or it was only at very low percentages. Instead, the experimental evidences better supported that the α-Gal/GalNAc unit is separated from the α-Fuc residue upon lectin binding, displaying slightly higher ϕ angles in comparison with those found in the free ligands. Besides the favorable corrections introduced on the CORCEMA-ST profile, the slight twist of the α-Gal/GalNAc glycosidic bond perfectly explained both the absence of a tr-NOE from Val351 for H4 α-Gal and the increased distance between H5 α-Gal and H1 α-Fuc upon binding, as observed

in the tr-NOESY and ROESY spectra of the bound and free ligands, respectively. Therefore, by virtue of these results, the binding geometry proposed for the histo blood group antigens A and B could be certainly considered as a robust 3D structure, jointly corroborated by experimental and computational approaches.

4.6. The 3D model of the complex

The data gathered so far by means of ligand-based approaches has permitted to propose a binding model that describes different features of the interaction between the DC-SIGN CRD and the histo blood group antigens A and B. In analogy to the existing data for the recognition of the Lewis X motif [38,59], the Fuc moiety is placed at the canonical binding site, establishing direct coordination bonds with the metal ion through OH3 and OH4 groups and additional vdW contacts with the surrounding amino acids. Certainly, the neighboring sugars also play secondary roles during the formation of these complexes, slightly tuning the affinity and the interaction dynamics in comparison with the free monosaccharides. Interestingly, the configuration of the Fuc glycosidic linkage in the blood group antigens (α1-2 instead of α1-3) leads these ligands to display a totally different sugar arrangement in its interaction with the lectin. Thus, in the Lewis X trisaccharide, the other non-reducing pyranose is placed into the secondary binding site formed by β-strands 2, 3 and 4, pointing its more aliphatic face towards Phe313, while the N-acetyl group from the central GlcNAc moiety accounts for the additional stabilizing hydrophobic contacts with Val351. In contrast, in the blood group antigens, this latter role involves the non-reducing Gal/GalNAc, while the opposite ligand regions are located near the secondary binding site, but without participating in the interaction.

At this point, some questions remained opened, especially concerning the differences detected by the protein-based NMR approach between both antigens. The non-reducing α-GalNAc moiety preferably displayed ϕ angles of ca. -60º, thereby enabling the formation of polar contacts with the protein through OH3 and/or OH4. In such an arrangement, the position 2 of the terminal α-GalNAc in the A antigen stays close to Val351 in a large percentage of the MD trajectories, suggesting the existence of an additional transient hydrophobic contact with the long loop. Strikingly, if this interaction exists, the affinity of the A antigen would be expectedly better. However, the experimental evidences pointed towards the opposite trend, since no tr-NOE effects were actually detected between the acetyl group and methyls from

Val351. Hence, the proposed model did not fully explain why these oligosaccharides behave differently in terms of dynamics and affinities, considering that the structural requirements for binding are equally fulfilled for both sugars. Apart, although the stabilizing effect arising from the non-reducing α-Gal is present in both cases, the affinity of the A antigen was slightly worse than that of the free monosaccharide (Fuc), as deduced from the binding isotherms built by ^{15}N-HSQC-based titrations. In principle, this fact could be attributed to the existence of alternative binding poses for the Fuc monosaccharide. Indeed, as highlighted before, the existence of several binding modes enhances the affinity of a given ligand (statistical rebinding).

Then, taking into account the promiscuous character exhibited by DC-SIGN on sugar recognition, the possible existence of additional binding modes was explored using T_2-filtered NMR experiments and the ^{19}F nucleus as probe. As precedent, the results from ^{15}N-HSQC-based titrations already supported this hypothesis in the case of the blood group B. As already detailed in the introductory section, the plasticity of the binding site is a common feature shared by C-type lectins and is the basis of their broad specificities. This ability not only enables the relatively easy accommodation of many different tridimensional structures, but also leads these receptors to recognize the same monosaccharide through alternative binding poses, as widely demonstrated for Man in DC-SIGN [60,73,79], and for Fuc between close lectins (MBP-C [75], Langerin [74] and DC-SIGN [59]) (Figure 4.52).

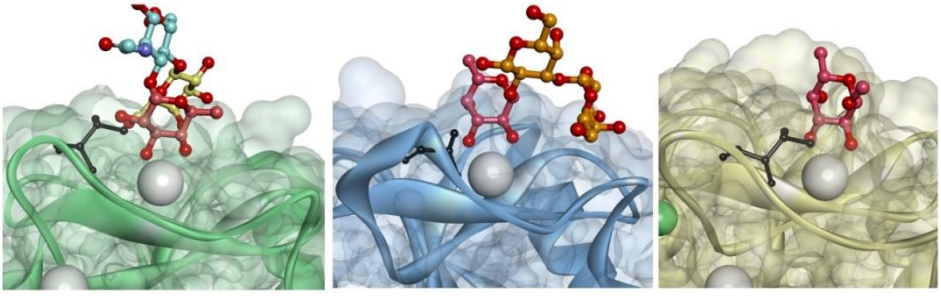


Figure 4.52. *Close view of the Ca^{2+}-dependent binding site of three related lectins harboring different Fuc-containing ligands. Left: DC-SIGN with LNFP III (PDB 1SL5); Middle: Langerin with the blood group B trisaccharide (PDB 3P5G); Right: MBP-C with α-OMe-Fuc (PDB 1RDI).*

Considering that the major interaction is driven by hydroxyls OH3 and OH4 of Fuc, and that Gal and Glc are very weak binders of DC-SIGN as well [36], the other existing alternatives would include:

- Through OH2-OH3 of Fuc.
- Through OH2-OH3 of the reducing Glc.
- Through OH2-OH3 or OH3-OH4 of the non-reducing α-Gal, for the B antigen.
- Through OH3-OH4 of the non-reducing α-GalNAc, for the A antigen.

The possibility of targeting the calcium ion by means of the inner Glc was immediately discarded, given that there was no way to place the tetrasaccharide at the primary binding site without generating important clashes involving the core β-Gal. The remaining options were evaluated by means of ^{19}F-T_2 relaxation experiments using a small library of fluorinated monosaccharides, including 2F-Fuc, 4F-OMe-Fuc, 2F-Gal and 4F-Gal as probes, and 3F-Man as negative control (Figure 4.53). This last sugar has been recently reported to completely lack the ability to bind to DC-SIGN [73]. In addition, 3F-Fuc was also included to directly compare its T_2 rate in the presence of the lectin with those of the 2F-Fuc and 4F-Fuc. In principle, the 3F-Fuc was neither expected to be recognized given that blockage of position 3 would preclude binding through either the OH3-OH4 or the OH2-OH3 pair. The chemical syntheses of 2F-Fuc [94-95], 2F-Gal [96], 4F-Gal [97] and 3F-Man [98] were already achieved and optimized several years ago and they are currently commercially available compounds. The 4F-OMe-Fuc was *de novo* synthesized and provided by the group of Prof. B. Linclau at the Faculty of Chemistry in the University of Southampton (UK). Finally, the 3F-Fuc synthesis was accomplished in collaboration with the same group, by a novel method developed during the short term stay that is detailed in chapter 6.

The relaxation experiments were performed at 298 K in the absence and the presence of 10 μM of lectin. To enhance the relaxation effects caused by the lectin and better evidence the binding events, the tetrameric extracellular domain of DC-SIGN (DC-SIGN ECD, 40 μM CRD) was employed rather than the CRD. The fluorine nucleus was individually assigned for each ligand through a simple proton-decoupled 1D-{^1H}-^{19}F (Figure 4.54). All the 12 fluorine signals (6 compounds, 2 anomers) were perfectly separated, and hence the time cost was reduced by only recording two T_2-filtered experiments with the mixture of monosaccharides, containing 0.8 mM of each. Obviously, although the ligand ratio would be 1:20 under these conditions, the real ratios were actually lower after the anomeric equilibrium had been reached for the free sugars (all of them except for 4F-OMe-Fuc). Nevertheless, the final ratios were still large enough (between 1:5 and 1:20) to record the CPMG spectra with good experimental sensitivity. Indeed, as stated in the *Methods* section, the

sensitivity of these experiments is fairly high and very weak interactions can be detected as well [99]. To avoid possible biasing arising from non-specific binding, a third control experiment was additionally recorded: a large excess of d_{12}-EDTA (20 mM) was added to sequester the available calcium in the buffer (4 mM) and disrupt the specific Ca^{2+}-dependent sugar recognition.

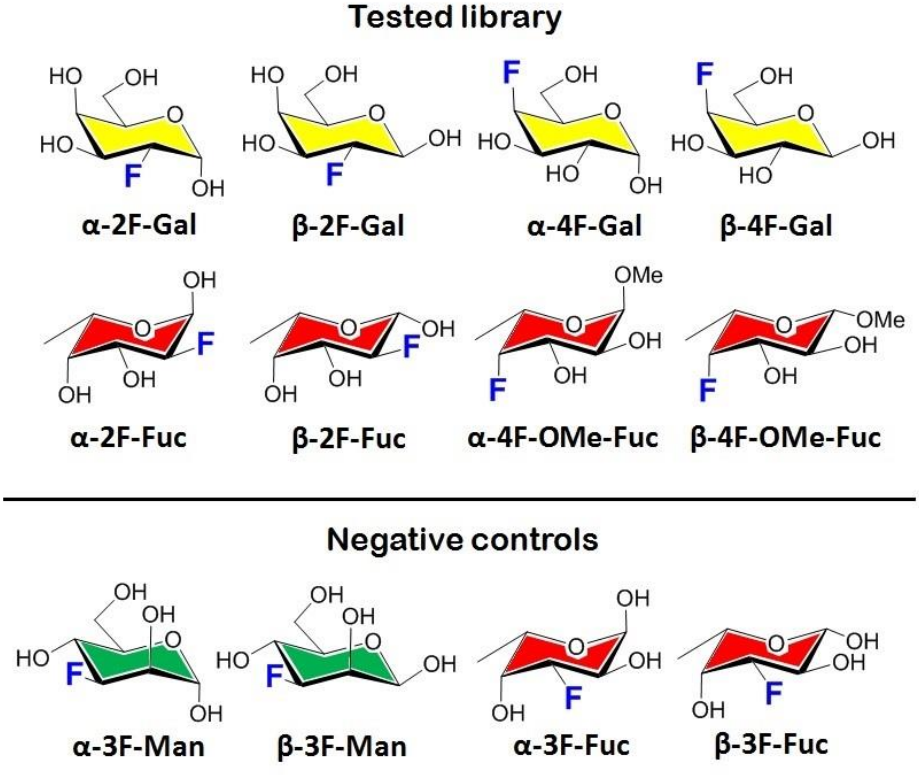


Figure 4.53. *Library of fluorinated monosaccharides used for performing ^{19}F-based T_2-filtered NMR experiments.*

The CPMG pulse sequence [100-101] was applied with a consensus τ = 2 ms and varying the number of echoes ("n") from 2 to 1300 (24 points). The decay of each ^{19}F signal was monitored for every "n" value, in the presence and in the absence of the lectin. From the individual graphs, the corresponding transverse relaxation times (T_2) were estimated by adjusting the data to Eq. 4 (Eq. 20 in *Methods*). To assess the binding ability of each ligand, the results were evaluated both graphically and also through the analysis of the percentage of T_2 decay, calculated by comparing the difference between both T_2 rates (free/bound) with the absolute $T_{2,free}$, as described by Eq. 5. The overall results are summarized in Figure 4.55.

$$I = I_0 \exp\left(-\frac{n2\tau}{T_{2,obs}}\right) \qquad (Eq.4)$$

$$\%\mathbf{T_2\ decrease} = \frac{T_{2,free} - T_{2,bound}}{T_{2,free}} \cdot 100 \qquad (Eq.5)$$

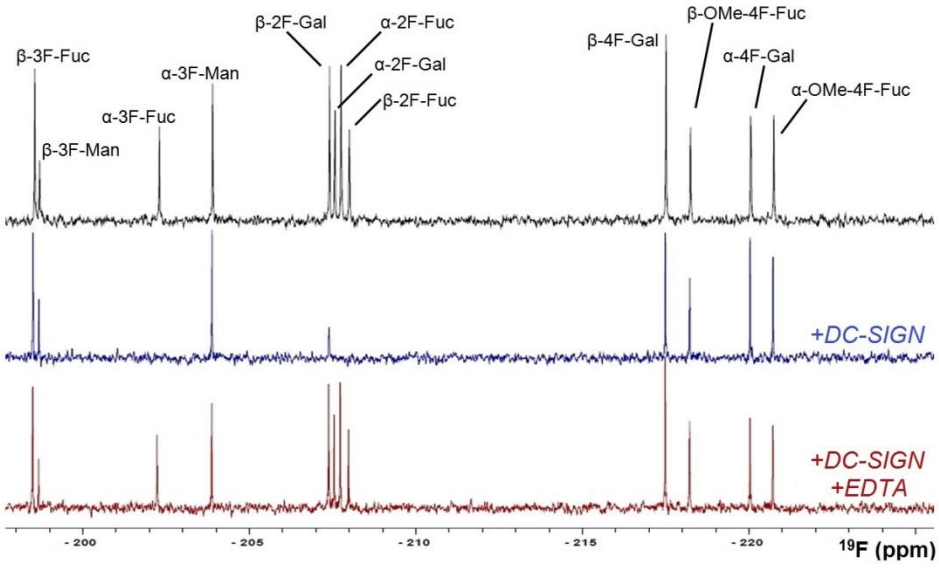
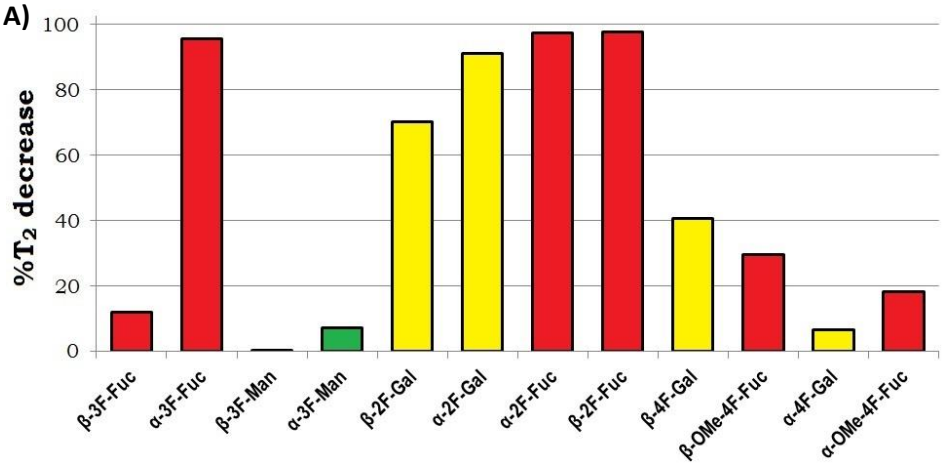


Figure 4.54. *Monodimensional ^{19}F-{^{1}H} spectra acquired at a consensus echo time of 1000 ms (n = 250) with the CPMG pulse sequence. The chosen echo time was optimal for the observation of large differences in the free/bound decays. Top: free ligands (0.8 mM). Middle: ligands + 40 µM CRD. Bottom: ligands + 40 µM CRD + 20 mM EDTA.*

As expected, neither the 3F-Man anomers nor the β-3F-Fuc could bind to the DC-SIGN ECD, thereby highlighting the essential role that position 3 plays in the recognition of both sugars by this lectin. Analogously, α-4F-Gal displayed no visible changes in its decay profile after adding DC-SIGN, thus totally discarding it as a binder. Instead, the T_2 decay was slightly faster for β-4F-Gal, as shown in Figure 4.55B, suggesting that the binding is negligible or extremely weak. The fact that β-4F-Gal preserves a residual binding character that does not exist for the α anomer is supported by the known ability of β-Gal to target the calcium site through its axial OH1 and equatorial OH2 when OH1 is available (reducing end), in an alternative pose called "non-physiological" binding [75]. Surprisingly, for both 4F-OMe-Fuc the scenario was rather similar to that described for the previous ligands, namely, the variations in the corresponding relaxation times were remarkably low and actually not meaningful. In retrospect, these observations differed from those works reporting that the binding of Fuc to other lectins is also feasible through

positions 2 and 3 [74-75]. Instead, regarding the results presented herein, it is clear that this OH2-OH3-based binding is much less favored in DC-SIGN. In contrast, the corresponding 2F-Fuc probes displayed a sharp drop in their ^{19}F signals upon protein addition, quickly disappearing only after 40 echo cycles due to the fast protein-mediated relaxation, and thereby unequivocally confirming these sugars as binders. Finally, both 2F-Gal anomers stood out as binders of DC-SIGN as well, as displayed in Figure 4.55A. Interestingly, the signal decay was fairly strong in comparison with that of 4F-OMe-Fuc, and they were actually among the clear strongest binders within the studied library just after both 2F-Fuc anomers. It has been reported that Gal can be weakly recognized by DC-SIGN [36], and the findings reported herein have highlighted that such a binding is mainly established by means of hydroxyl groups OH3 and OH4, as shown for Fuc. Moreover, the recognition of the α anomer was clearly more favored than that of the β anomer. Finally, it is specially worth noting the particular case of the α-3F-Fuc. Certainly, it was reported to produce a noticeable decay in the presence of DC-SIGN. Taking into account the completely opposite scenario found for the β-anomer, the anomeric position is very likely to be involved in the binding together with nearby OH2. Indeed, the stereochemistry of OH1 and OH2 is exactly equivalent to that of OH3 and OH4, hence this observation could be reasonably justified as well as it also supports the slightly better affinity of Fuc found in this work in comparison with its corresponding methoxy-L-glycosides [36]. To note, in all cases, and especially for those fluorinated sugars that were clearly recognized by the ECD, the further addition of EDTA led to almost complete signal recovery, what eventually proved that for all the probes the binding was always taking place in a specific manner, at the Ca^{2+}-dependent primary binding site.



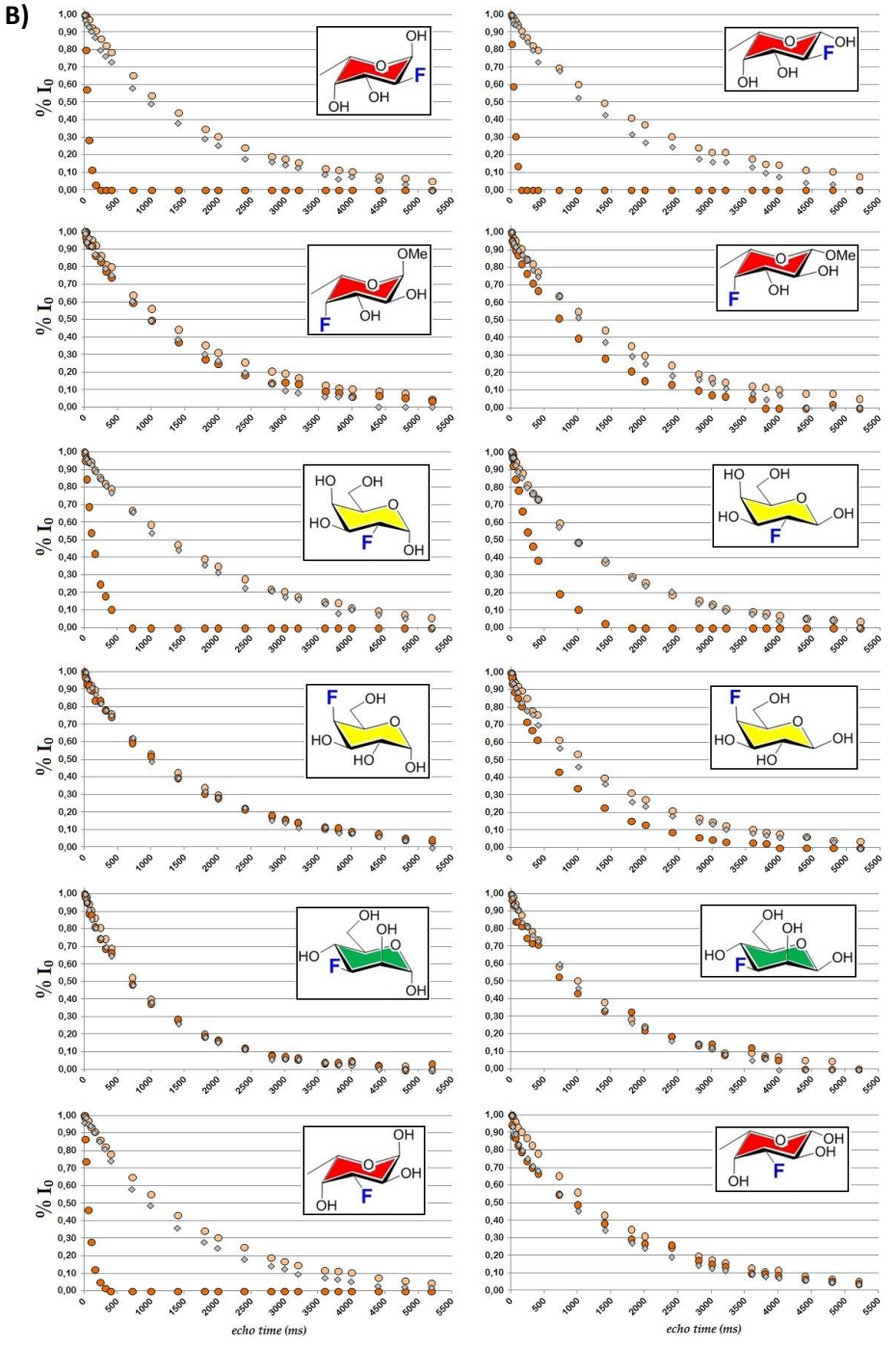


Figure 4.55. *A) Percentage of T_2 decrease calculated by means of Eq. 5. Each T_2 value was individually obtained by fitting data shown in B) to Eq. 4. B) Signal decay curves reported for every single ligand. The intensity values are displayed in relative scale, giving the value 1.00 to the intensity measured at the first point (n = 2). Each graph includes three curves: the free sugar (light orange circles), the sugar in the presence of the lectin (dark orange circles) and after adding 20 mM of EDTA (grey diamonds).*

Overall, these results point out that the preferential interaction with the DC-SIGN ECD is established by the Fuc moiety, chelating the calcium ion chiefly through hydroxyls at positions 3 and 4. Fittingly, this is the major mode in which DC-SIGN interacts with both blood group antigens A and B, as demonstrated in the previous sections. Regarding the other available residues, the existence of a secondary binding pose for the blood group B would be now supported by the additional participation of the non-reducing α-Gal, which would target the binding site through OH3 and OH4. In the same line, the contribution of any other binding mode involving calcium targeting through positions 2 and 3 is probably negligible. Taking all these observations together, the Fuc moiety would only enable the interaction of the blood group antigens by means of a unique binding mode. Additionally, the B antigen, but not the A, would have the ability to rebind to the DC-SIGN CRD through its terminal α-Gal, thereby explaining its slightly better affinity and more complex binding isotherms. Of course, to fully confirm this hypothesis, the interaction through the GalNAc moiety in the case of the A antigen should also be discarded. Since a fluorinated GalNAc moiety could not be included in the ^{19}F-T_2 experiments, the normal GalNAc monosaccharide was used to perform STD experiments with DC-SIGN. Also, Gal was included in the study with the purpose of gathering data for proposing a feasible binding pose, as done for the Fuc-based recognition models. As the Gal-mediated interaction presumably takes place at a low percentage, this information cannot be directly extracted from an STD spectrum with the B antigen, in which the STD effects are essentially contributed by the major binding structure.

Following the same experimental methodology, STD spectra were performed with a fixed total ligand concentration of 4.2 mM (in D_2O) and a protein/ligand ratio of 1:70 (60 μM of CRD). The saturation was applied for 2 seconds at the aliphatic region (0.76 ppm for Gal and GalNAc; 0.60 ppm for Fuc). In this case, according to the structure of the blood groups, only the methyl α-glycosides were selected and employed (α-OMe-Gal, α-OMe-GalNAc, α-OMe-Fuc) in order to simplify the analyses and minimize signal overlapping issues. Unfortunately, both anomeric H1 from α-OMe-Gal and α-OMe-GalNAc resonated too close to the signal water, and a consensus temperature could not be determined to observe them. Thus, the spectra were recorded at 288 K, which was the only temperature that allowed measuring at least the H1 from α-OMe-Fuc. Although the affinities of these ligands were in the same range or lower than those of the tetrasaccharides, the exchange regimes were expectedly faster, as already proved for Fuc by ^{15}N-HSQC-based titrations. Then, the higher inefficiency of the saturation transference resulted in a

notorious reduction of the experiment sensitivity, that could be partially compensated by two ways: applying a spin-lock filter to suppress the water signal (without deeply affecting H1 Fuc) and recording an additional STD of the free protein (receptor blank), that was subtracted from the final STDs to efficiently remove the protein background. These final spectra are commonly named STDD (Saturation Transfer Difference Difference) (Figure 4.56).

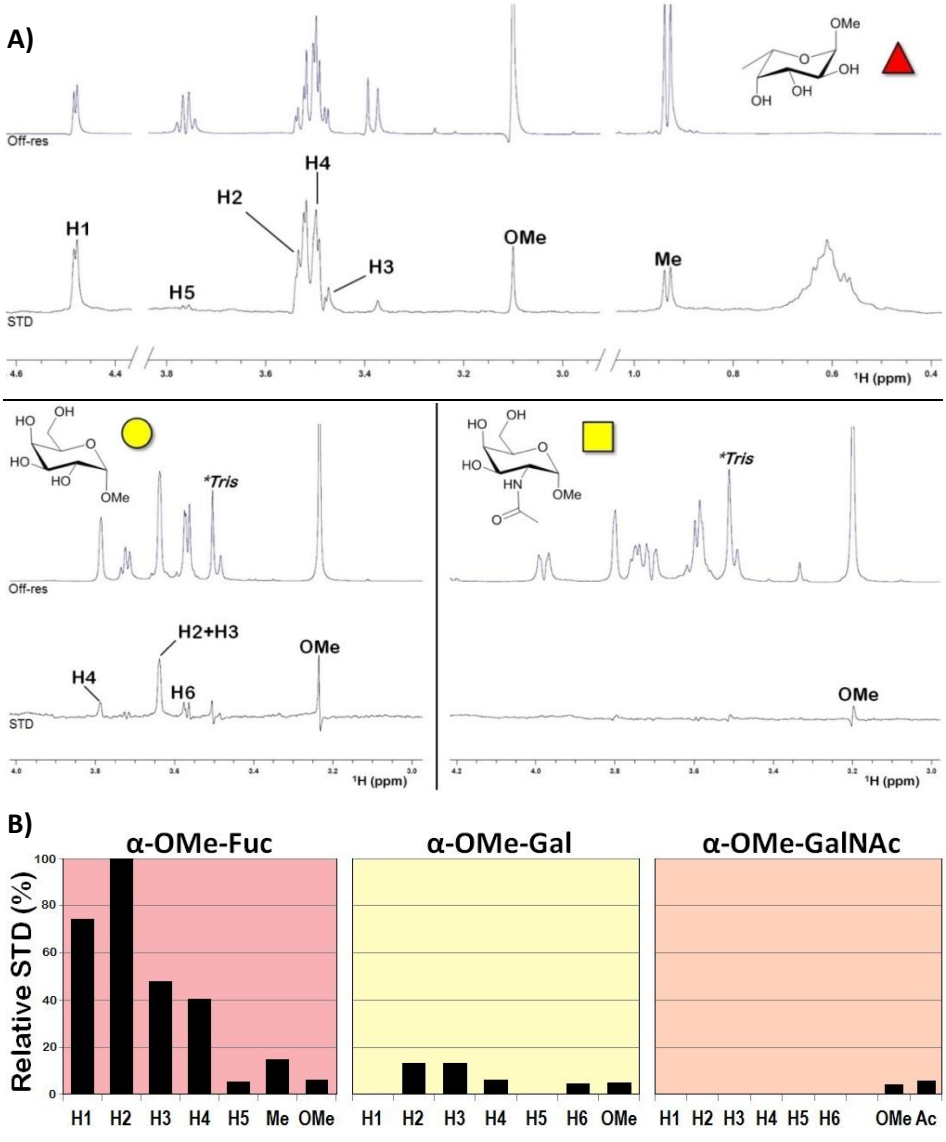


Figure 4.56. *A) Identification of the visible STDs reported for each individual monosaccharide. B) Global analysis of the relative STD intensities obtained for the three ligands, taking the Fuc H2 proton as the overall reference nucleus, with a relative intensity of 100%.*

The STDD of the α-OMe-Fuc essentially yielded the same relative STD effects as those observed for the blood groups (Figure 4.56). The H2 proton clearly displayed the strongest STD (STDrel = 100%), followed by H1. In addition, weaker STDs could be also noticed for H3 and H4, ranging around 40-50% in relative terms, whereas the contribution of the remaining protons was fairly lower (below 15%). For α-OMe-Gal, the strongest STD showed a relative intensity of ca. 14% as compared with H2 Fuc, evidencing the lower affinity of this monosaccharide. To note, that effect was determined as the sum of protons H2 and H3. The other nuclei only provided very weak or residual STDs that were still measurable though, except for H5. Besides H2 and H3, saturation was also noticeable for H4. This proton showed the second strongest relative STD, near to 50% if compared to H2/H3 Gal, while only 7% if compared to H2 Fuc.

Finally, the STDD of the α-OMe-GalNAc provided no STD peaks at all. Even subtracting the receptor blank spectrum, only residual STD effects could be measured for the N-acetyl and the O-methyl groups (lower than 5%). Nevertheless, these ones were better interpreted as artifacts derived from the specific pulse sequence applied, as they did not have any sense regarding the monosaccharide structure and the possible ligand arrangements.

In summary, these results further assessed that both blood groups A and B target the canonical calcium site of DC-SIGN primarily through the hydroxyl groups OH3 and OH4 of their respective Fuc residues and explain why they are differentiated in terms of affinities and dynamic behavior. Accordingly, a minor binding mode takes place only for the B antigen, which is established through the non-reducing α-Gal residue and entails an additional stabilization of the complex formed by this antigen.

To further demonstrate whether this alternative Gal-driven binding mode was still feasible when considering a larger oligosaccharide, additional NMR experiments were carried out for the lectin in the presence of a related antigen: the Galili type VI trisaccharide. In structural terms, this antigen is equivalent to the blood group B tetrasaccharide type VI, but it lacks the terminal α1-2 Fuc and the only non-reducing sugar is therefore the α1-3-linked Gal. Following the reasoning proposed above, the binding of this antigen to DC-SIGN should be exclusively possible through the terminal α-Gal. The specificity of this interaction was simultaneously proved by ^{15}N-HSQC-based titrations and STD experiments. Of course, for the latter ones, the proton chemical shifts of the ligand had to be assigned. However, this task

resulted fairly trivial and fast given that in the ^{13}C-HSQC spectrum most of the resonances were relatively located at the same positions than those observed for the analogous sugars in the blood groups. Thus, the identification of each proton via 2D-TOCSY spectra was quickly accomplished (Figure 4.57).

The perturbations on the DC-SIGN CRD amide crosspeaks were evaluated by ^{15}N-HSQC, using a 100 µM ^{15}N-labeled CRD sample. To note, due to the markedly poor interaction, the saturation plateau could not be reached and the affinity was not calculated either. In spite of the weak binding, the resulting spectra at 50-70 eq of ligand already showed noticeable shifts on several amino acids. A thorough analysis of the corresponding peaks revealed that about 11-13 residues were moderately shifted in the presence of the trisaccharide, and specifically four of them were remarkably perturbed (Figure 4.58 and 4.59A). These four residues were located very close to the calcium ion, both at the long loop (N349, V351) and at the opposite β_4 strand (D366, K368). From an overall perspective, the CSP distribution pattern for Galili was fairly similar to these previously evidenced for the Fuc-containing and Man-containing oligosaccharides. Indeed, a graphic analysis using the CRD structure quickly evidenced that the most relevant perturbations were all distributed along the protein region surrounding the calcium site 2 (Figure 4.59B). Hence, the interaction was likewise deduced to take place in a specific fashion, as reported for the blood groups, while the affinity was expectedly reduced, as inferred from the large amounts of ligands that were needed to achieve observable CSPs.

Interestingly, the Val351 amide group was actually rather affected in comparison with the blood groups, even more that the neighboring N350. The same difference was also noticed for K368, which was previously observed to undergo large peak displacements exclusively in the presence of the Man epitope, but not with Fuc. Moreover, the only clear long-range perturbations were reported for S310 and T314, whereas the nearby F313 residue remained unperturbed. From a dynamic point of view, and similar to the titration with Fuc, the changes in the peak positions were herein accompanied with minor broadening effects, which were indicative that the binding process was predominantly occurring within a fast exchange regime. In this line, and supposing a k_{on} value of the same order for all these ligands (Galili and blood groups), the lower affinity would be essentially imposed by a larger k_{off} rate. Obviously, the approximate values could not be estimated since the empirical equation used for the blood groups was only valid for cases of intermediate exchange.

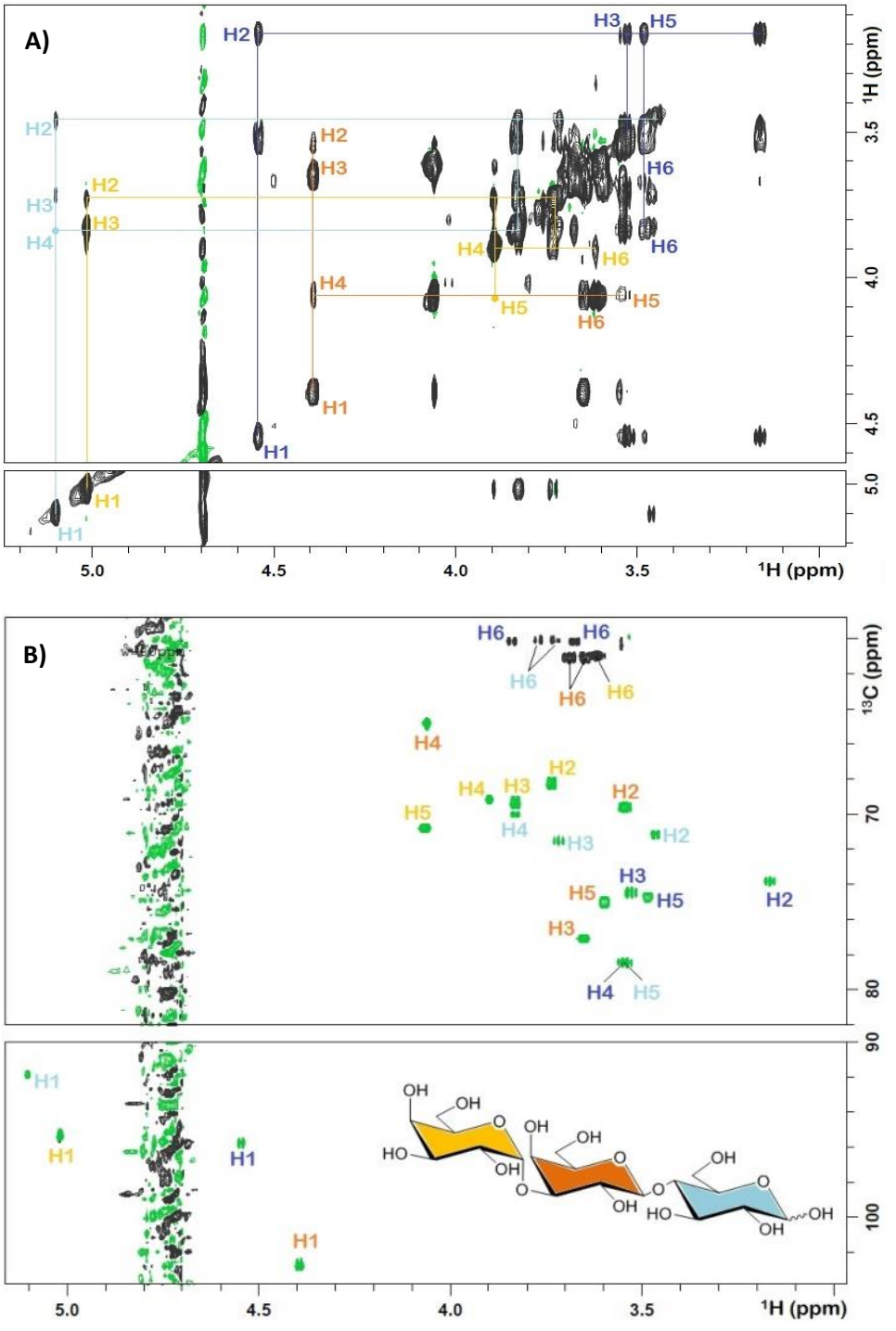


Figure 4.57. A) *2D-TOCSY correlations established for each independent sugar in the Galili antigen. The individual spin systems are detailed in different colors. B)* ^{13}*C-HSQC full assignment of the studied trisaccharide. Color code: Blue: β-Glc; light blue: α-Glc; orange: β-Gal; light orange: α-Gal.*

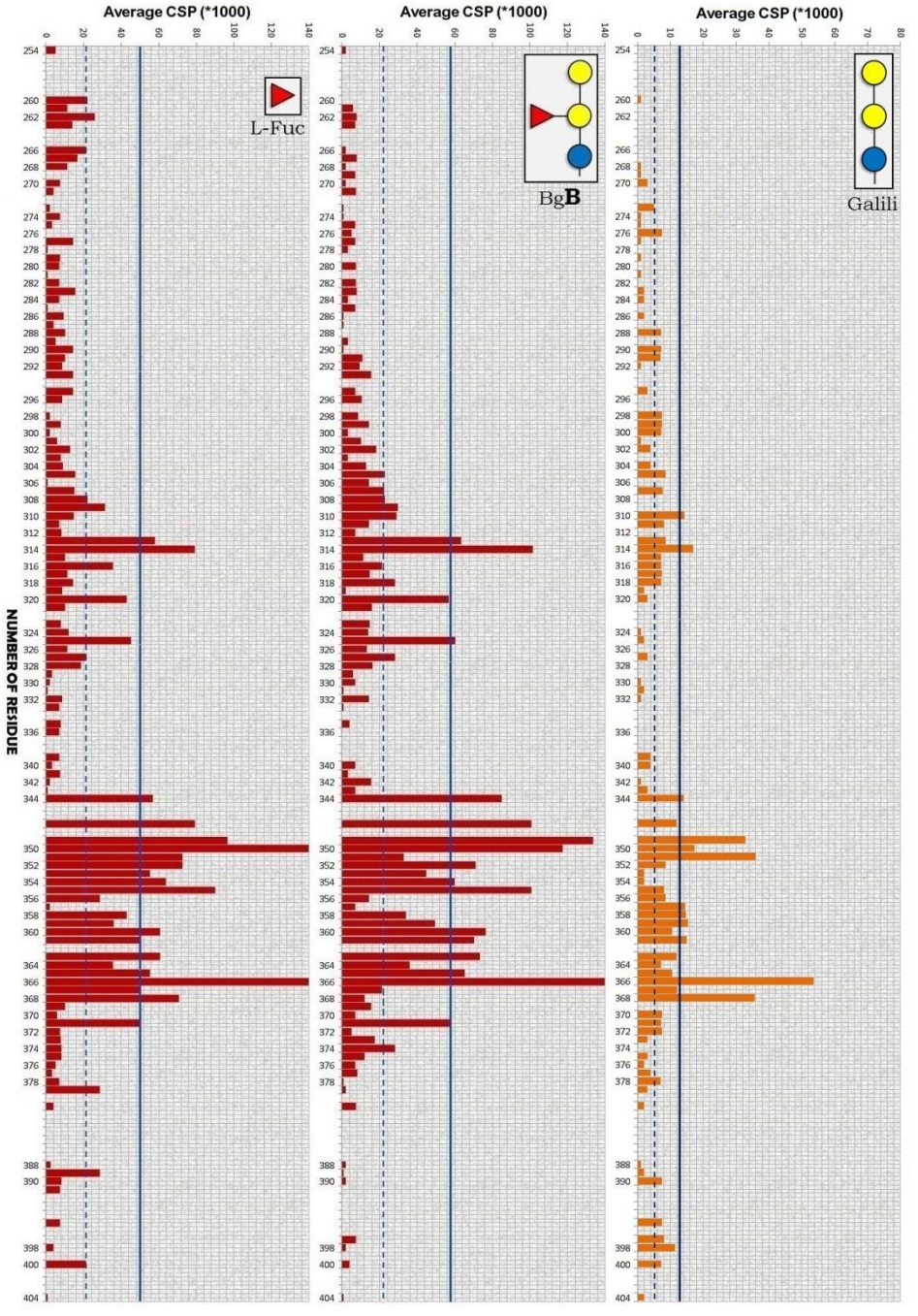


Figura 4.58. *Average CSPs measured for the Galili epitope (60 eq, right, in orange), compared with those reported for a blood group antigen (140 eq, middle) and the Fuc monosaccharide (140 eq, left). All data were collected at 310 K.*

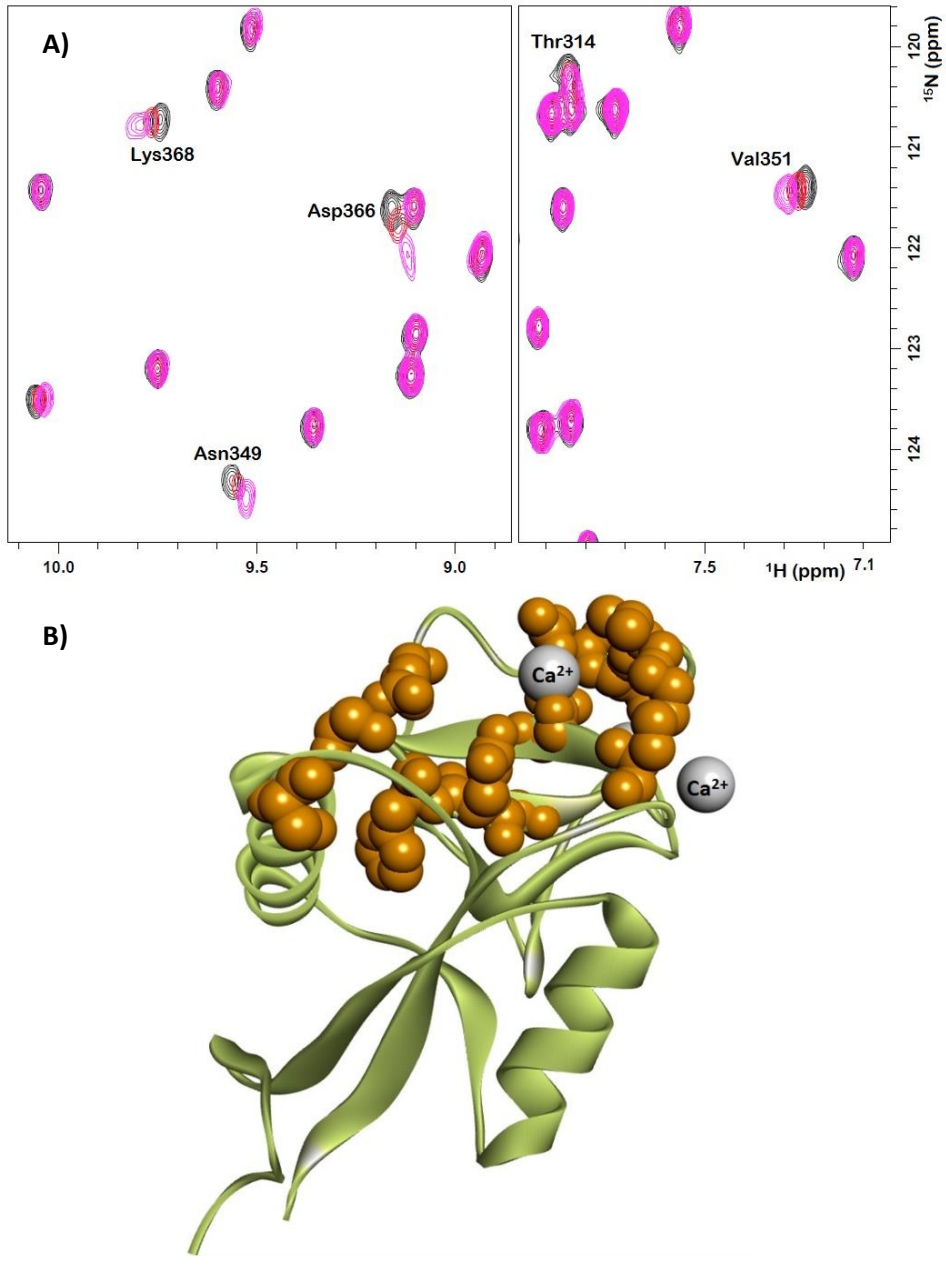


Figure 4.59. *A) Two sections of the ^{15}N-HSQC spectra of DC-SIGN recorded in the presence of the Galili epitope. Some residues significantly perturbed are labeled. Color code: black: free CRD; red: 36 eq of Galili; pink: 72 eq of Galili. B) The DC-SIGN CRD showing those amino acids that experienced CSPs above the cutoff value. Cutoff CSP = μ + σ = 0.0131 ppm.*

The ^{15}N-HSQC titrations unequivocally proved the existence of a specific interaction between the Galili trisaccharide and the DC-SIGN CRD, thereby demonstrating that the recognition of this scaffold is weak but still possible in

the absence of Fuc. Subsequently, these studies were complemented with STD measurements, applying the same experimental conditions described above (T = 288K, 1:70 protein/ligand, aliphatic saturation (2 sec.)). After subtracting the protein blank, the resulting STDD allowed to distinguishing a clear set of weak STDs arising from the studied trisaccharide (Figure 4.60). Those measurable NMR signals ranged between 0,5% and 1% of absolute STD intensity and essentially came from the two Gal moieties. The non-reducing α-Gal yielded the predominant STDs at H1, H2 and H3, with intensities within the 0,8-0,9% range. In relative terms, these intensities were about 25% if compared with H2 of the α-OMe-Fuc monosaccharide, and therefore slightly higher than those measured for α-OMe-Gal (Figure 4.61). For the central β-Gal, some saturated protons could be identified as well, namely H3, H4, H5 and H6. Their respective contributions were substantially lower than those of the terminal α-Gal. Regarding the remaining reducing Glc, their signals were barely detectable and so attenuated that it was unworkable to accurately measure any STD intensity (<<0.5%).

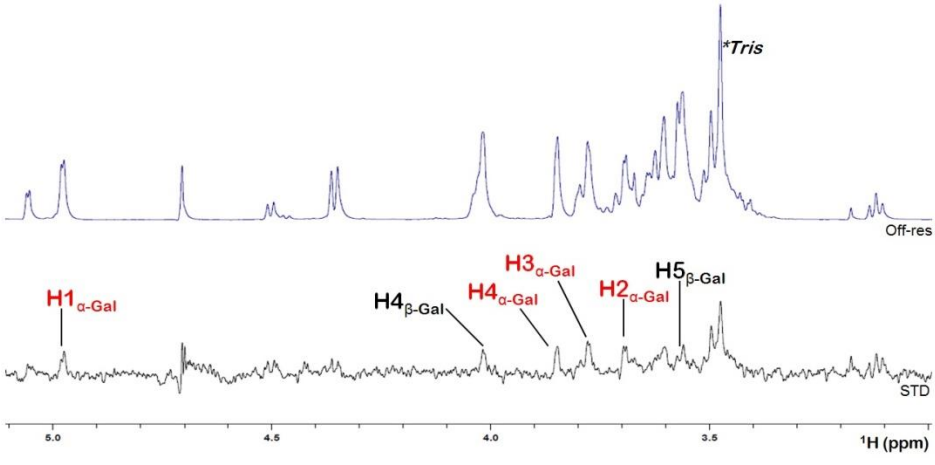


Figure 4.60. *STDD profile obtained for the Galili trisaccharide at 288 K in the presence of 60 μM of CRD (70 times lower). The saturation pulse was applied at 0.76 ppm (aliphatic region) for 2 seconds. The STDs provided by the non-reducing α-Gal are indicated in red.*

Unfortunately, the acquisition of a Tr-NOESY in the case of the Galili epitope was not possible given the low affinity of the ligand. Then, the STD and the ^{19}F-T_2 data constituted the unique piece of information from the ligand perspective to unveil the binding mode. In this scenario, a manual docking protocol was attempted to explain the STD observations and to deduce the structure of the lectin-sugar complex by Molecular Dynamics.

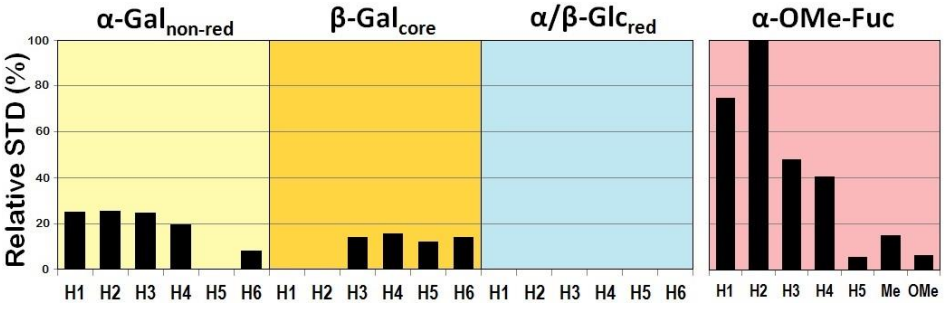


Figure 4.61. *Relative STDs calculated for the three residues of the Galili epitope, taking the H2 proton from the α-OMe-Fuc monosaccharide as reference (right, separated in the red box).*

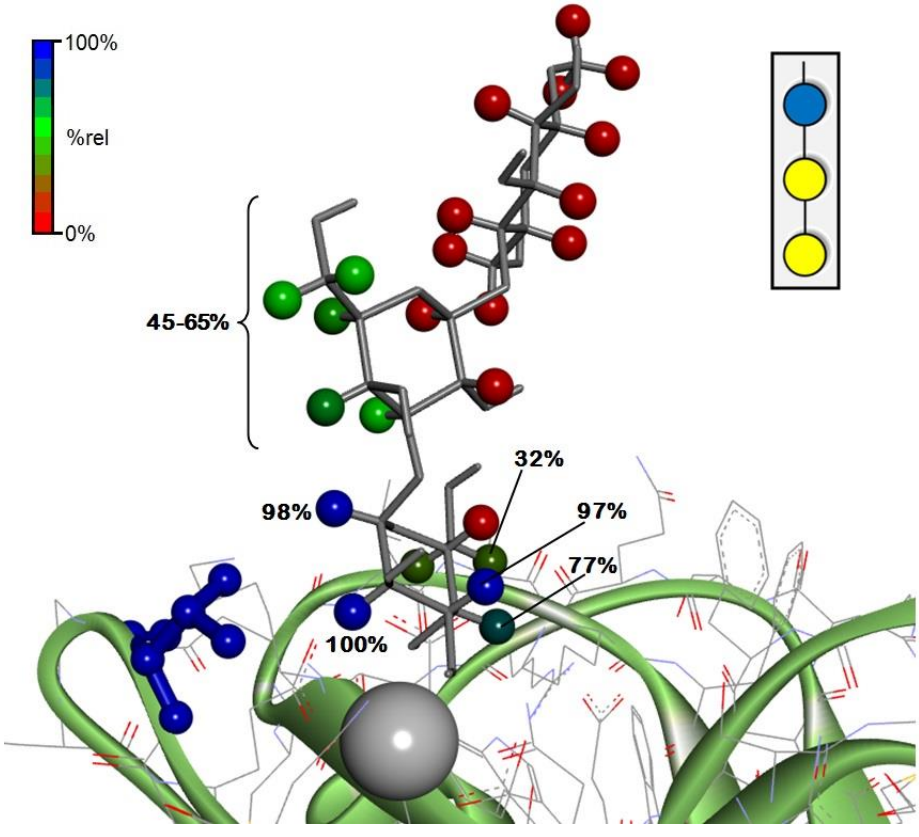


Figure 4.62. *Proposed binding model for the interaction between the DC-SIGN CRD and the Galili antigen type VI. All the non-polar protons are represented as colored spheres. The color depends on the percentage of relative STD, being H2 α-Gal the reference nucleus (100%, blue). Several notorious STDs are directly indicated. The Val351 sidechain is also depicted as blue tight sticks.*

From the relaxation studies, the binding mediated by the Gal pyranose is meant to occur through hydroxyl groups at positions 3 and 4. At this point, a

further complication appeared related to the sugar arrangement: since no interaction complexes between DC-SIGN and Gal have been described from a molecular perspective so far, the binding pose of the Gal ring had to be adapted starting from another ligand. Then, the PDB 1SL5 was used again for this task. After modeling the trisaccharide structure with the GLYCAM online server, O3 Gal and O4 Gal were suitably placed to provide the coordination bonds with the metal ion, using the Fuc ring from LNFP III as template. In this arrangement, the most hydrophobic face of the pyranose ring is pointing towards the Val351 sidechain, in agreement with the observed STD (Figure 4.62). Noteworthy, as the sugar symmetry is different for Gal (D-sugar), its orientation in this model was just the opposite to that of the L-Fuc ring (rotated 180º). In the inverted pose, α-Gal H1 and H2 (which provided the highest STD) would be placed far away from Val351, whereas α-Gal H5 (whose STD was negligible) would be the closest proton to the Val sidechain. Therefore, that binding configuration was discarded according to the available experimental evidences.

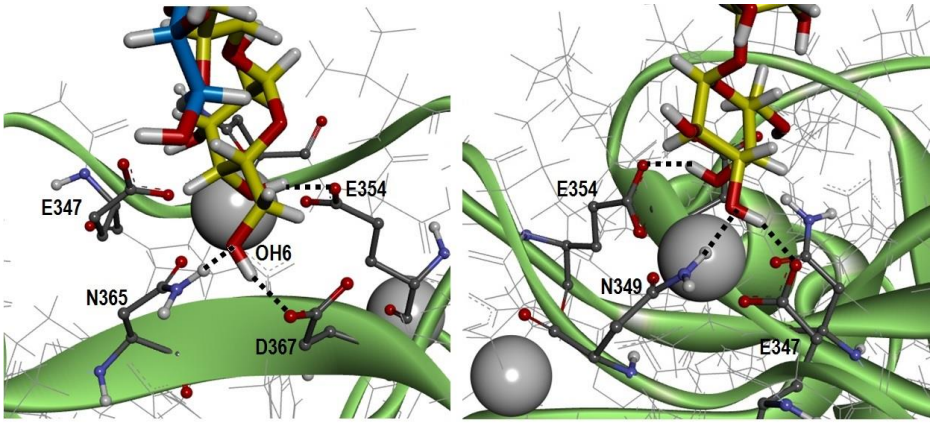


Figure 4.63. *Front and back views of the primary binding site displaying the intermolecular interactions established between the non-reducing Gal moiety of Galili and the surrounding polar amino acids (highlighted as tighter lines). Although not shown in this particular pose, the contact between OH4 and N365 is possible in other frames of the MD trajectory. The coordination bonds are not represented*.

The stability of the proposed model was eventually assessed by 200 ns of MD simulation. Noteworthy, the trisaccharide remained attached to the binding site along the entire simulation, with the non-reducing Gal establishing stable and fixed hydrogen bonds with several surrounding amino acids that coordinated the metal ion as well. A more detailed inspection revealed that these residues were the same as those involved in Fuc binding: E347, N349, E354, N365 and D367 (Figure 4.63). Besides the participation of hydroxyls OH3

and OH4, position 6 of the terminal Gal also seemed to be fairly important for the hydrogen bond network, instead of OH2 as reported in Fuc. In fact, the presence of the OH6 group was mandatory along the simulation to target the polar sidechain of D367, which lay too far from OH3 and OH4. Similarly, N365 simultaneously participated in the interaction with OH4 and, in a lesser extent, with OH6 as well.

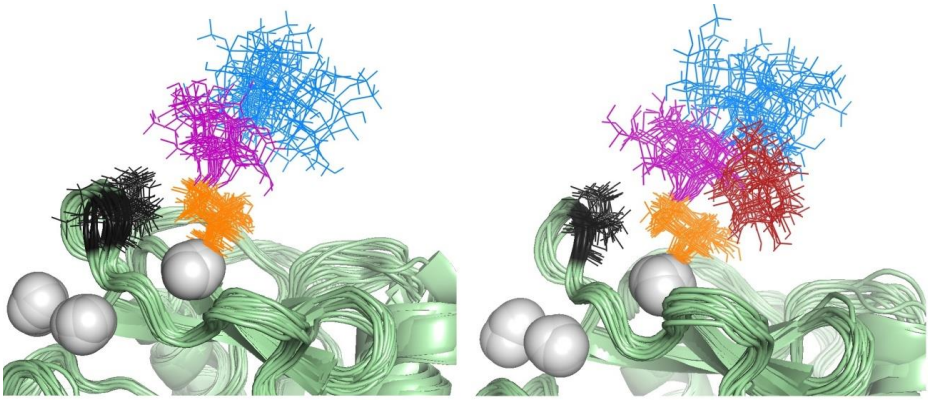


Figure 4.64. *Ligand trajectories registered for the Galili antigen (left) and the blood group B (right) along 200 ns of MD simulation, represented as bundles containing 20 representative frames. In both cases, the oligosaccharides are attached to the recognition site through the Gal residue (orange). Color code: blue: β-Glc; purple: β-Gal; red: α-Fuc; orange: α-Gal. The white spheres represent the calcium ions. The residue Val351 is detailed in black.*

Also noting, the flexibility of the trisaccharide was considerably higher than that of the blood groups, especially concerning the Galα(1-3)Gal linkage. Consequently, the non-reducing Gal was the unique sugar residue placed at a relative fixed position, imposed by the two coordination bonds (Figure 4.64). On the contrary, the other two residues were along a wide space region, always away from the protein surface. Specifically, the average values for the ϕ torsions agreed again with the presence of *exo-syn* conformations (Figure 4.65), although the standard deviations were above 15º, meaning that the most preferred ϕ values were distributed within a broad range of 30 degrees or higher. Interestingly, two populations were detected for ψ in the Galα(1-3)Gal bond, centered around ψ = -40º (the major) and ψ = +35º (the minor). From the point of view of the ligand recognition, the existence of these two populations did not affect the binding, since no relevant vdW contacts with the lectin surface were provided by the central Gal moiety in any case.

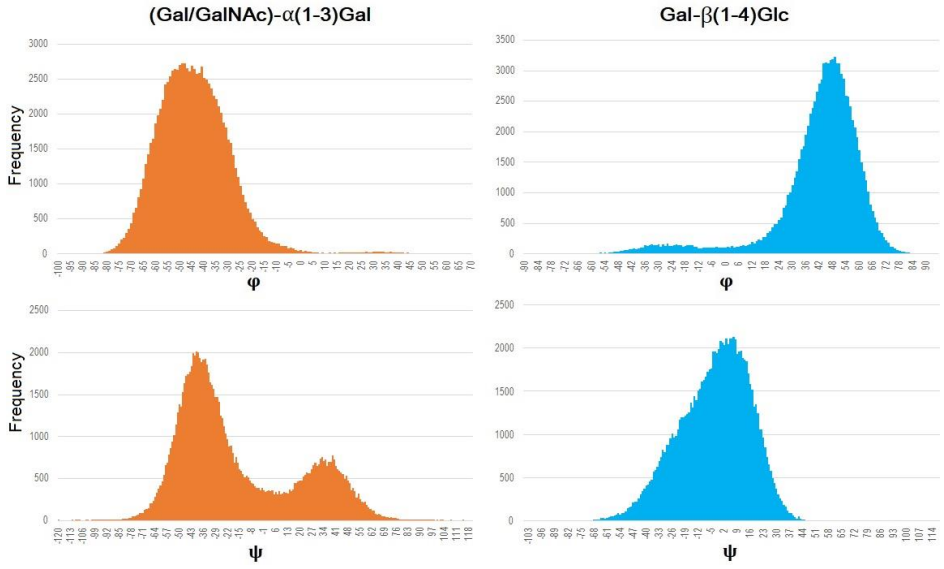


Figure 4.65. *Distribution of the torsional angles φ (top graphs) and ψ (bottom graphs) observed for each glycosidic linkage (orange/blue, indicated on the top), extracted from 100000 points of the MD run. The vertical axis indicates the frequency (number of frames) of every angle value.*

Finally, as a complementary test, one additional MD simulation was performed by placing the blood group B antigen at the binding site, but in this case anchoring it through the non-reducing Gal pyranose. Nicely, the tetrassacharide stayed at the binding site over the 200 ns of the MD run, thereby assessing the stability of this alternative binding mode. Certainly, the mobility of the ligand was reduced in comparison with the Galili epitope, as expected given the lower structural flexibility exhibited by the tetrasaccharides (Figure 4.64). However, the overall mobility was still high if compared to that noticed during the Fuc-mediated binding. In principle, the two coordination bonds with the calcium ion substantially fixed the orientation of the interacting Gal moiety, although the exact sugar position was noticed to remain rather malleable. Obviously, the presence of additional structural elements contacting the receptor surface, like bulky substituents (other sugars), is expected to limit the ligand positioning when coordinating the metal center. Thus, a reasonable explanation for the case of study could arise from the recognition roles played by the neighboring sugars: the major binding pose is partially defined by the arrangement of the non-reducing Gal/GalNAc residue, which provides stable contacts with the protein, while in the Gal-driven interaction model the recognition exclusively involves the sugar that establishes the coordination bonds. Indeed, the Fuc residue did not

provide any relevant intermolecular interactions with the nearby protein sidechains.

In summary, the recognition of the blood groups by the DC-SIGN CRD has been demonstrated to chiefly take place through the α(1-2)-linked Fuc residue. The direct coordination at the calcium site 2 is driven by the oxygen atoms situated in the adjacent positions 3 and 4 of the pyranose ring. Additionally, the other non-reducing end (Gal/GalNAc) is also involved in van der Waals contacts with neighboring protein amino acids, thus conferring these ligands slightly improved affinities. From a dynamic point of view, the lectin targeting occurs in an intermediate exchange regime in the chemical shift time scale, differing from the corresponding monosaccharides (Fuc, Man) that exhibit faster k_{off} rates. Importantly, the A antigen displays a unique binding mode, which is also major for the B antigen. Moreover, in the latter case a secondary binding mode also takes place, as evidenced by NMR. To highlight, the major geometry deduced herein (Figure 4.66) is consistent with the crucial role that Val351 plays in molecular recognition processes involving fucosylated structures [38,59,80]. In this particular case, both non-reducing sugars (Fuc and Gal) simultaneously account for the positive hydrophobic stabilization.

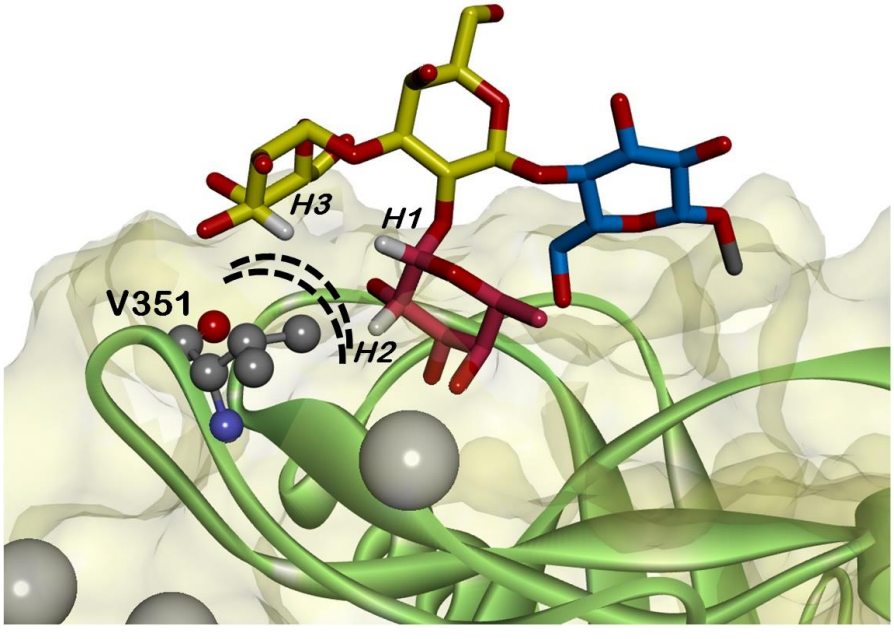


Figure 4.66. *Final binding model proposed for the histo blood group antigen B, detailing the protons that directly make contacts with the sidechain of Val351 at the long loop.*

4.7. References

1. R. D. Cummings and M. E. Etzler, *Essentials of Glycobiology*, eds. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler. Cold Spring Harbor Laboratory Press (New York), 2nd edn, 2009, **ch. 6**.
2. A. Varki. Biological roles of glycans. *Glycobiology*, (2017), **27**, 3–49.
3. C. Reily, T. J. Stewart, M. B. Renfrow and J. Novak. Glycosylation in health and disease. *Nat. Rev. Nephrol.*, (2019), **15**, 346–366.
4. P. Valverde, A. Ardá, N. C. Reichardt, J. Jiménez-Barbero and A. Gimeno. Glycans in drug discovery. *Medchemcomm,* (2019), **10**, 1678-1691.
5. P. J. Winterburn and C. F. Phelps. The significance of glycosylated proteins. *Nature,* (1972), **236**, 147–151.
6. H. J. Gabius, S. André, J. Jiménez-Barbero, A. Romero and D. Solís. From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem. Sci.,* (2011), **36**, 298-313.
7. H. Kaltner, J. Abad-Rodríguez, A. P. Corfield, J. Kopitz and H. J. Gabius. The sugar code: letters and vocabulary, writers, editors and readers and biosignificance of functional glycan-lectin pairing. *Biochem. J.,* (2019), **476**, 2623-2655.
8. A. Adibekian, P. Stallforth, M.-L. Hecht, D. B. Werz, P. Gagneux and P. H. Seeberger. Comparative bioinformatics analysis of the mammalian and bacterial glycomes. *Chem. Sci.,* (2011), **2**, 337–344.
9. D. H. Williams, E. Stephens, D. P. O`Brien and M. Zhou. Understanding noncovalent interactions: ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes. *Angew. Chem. Int. Ed.,* (2004), **43**, 6596-6616.
10. J. L. Asensio, A. Ardá, F. J. Cañada and J. Jiménez-Barbero. Carbohydrate-aromatic interactions. *Acc. Chem. Res.,* (2013), **46**, 946-954.
11. F. P. C. Binder, K. Lemme, R. C. Preston and B. Ernst. Sialyl Lewis(x): a "pre-organized water oligomer"? *Angew. Chem. Int. Ed.,* (2012), **51**, 7327-7331.
12. H. C. Siebert, J. Rosen, K. Seyrek, H. Kaltner, S. André, N. V. Bovin, P.G. Nyholm, F. Sinowatz and H. J. Gabius. Alpha2,3/alpha2,6-Sialylation of N-glycans: non-synonymous signals with marked developmental regulation in bovine reproductive tracts. *Biochimie.,* (2006), **88**, 399-410.
13. S. C. Purcell and K. Godula. Synthetic glycoscapes: addressing the structural and functional complexity of the glycocalyx. *Interface Focus,* (2019), **9**, 20180080.
14. H. S. Benett. Morphological aspects of extracellular polysaccharides. *J. Histochem. Cytochem.,* (1963), **11**, 14–23.
15. T. K. Dam and F. C. Brewer. Maintenance of cell surface glycan density by lectin–glycan interactions: A homeostatic and innate immune regulatory mechanism. *Glycobiology,* (2010), **20**, 1061–1064.
16. T. K. Dam and F. C. Brewer. Lectins as pattern recognition molecules: the effects of epitope density in innate immunity. *Glycobiology,* (2009), **20**, 270–279.

17. E. Maverakis, K. Kim, M. Shimoda, M. E. Gershwin, F. Patel, R. Wilken, S. Raychaudhuri, L. R. Ruhaak and C. B. Lebrilla. Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: a critical review, *J. Autoimmun.,* (2015), **57**, 1–13.
18. A. Ardá and J. Jiménez-Barbero. The recognition of glycans by protein receptors. Insights from NMR spectroscopy. *Chem. Commun. (Camb).,* (2018), **54**, 4761-4769.
19. S. M. Muthana, C. T. Campbell and J. C. Gildersleeve. Modifications of glycans: biological significance and therapeutic opportunities. *ACS Chem. Biol.,* (2012), **7**, 31-43.
20. H. Schachter and S. Roseman. Mammalian Glycosyltransferases. In: *The Biochemistry of Glycoproteins and Proteoglycans*, eds. W. J. Lennarz. Springer (Boston). 1st edn, 1980.
21. J. Roth. Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. *Chem. Rev.,* (2002), **102**, 285–303.
22. M. Aebi, R. Bernasconi, S. Clerc and M. Molinari. N-Glycan structures: recognition and processing in the ER. *Trends Biochem. Sci.,* (2010), **35**, 74–82.
23. I. Brockhausen and H. Schachter. Glycosyltransferases involved in N- and O-glycan biosynthesis. In *Glycosciences: Status and Perspectives*, eds. H.-J. Gabius and S. Gabius. Chapman & Hall (Weinheim), 1st edn, 1997, pp. 79-113.
24. P. R. Crocker, J. C. Paulson and A. Varki. Siglecs and their roles in the immune system. *Nat. Rev. Immunol.,* (2007), **7**, 255-256.
25. W. Reutter, R. Stäsche, P. Stehling and O. Baum. The biology of sialic acids: insights into their structure, metabolism and function in particular during viral infection. In *Glycosciences: Status and Perspectives*, eds. H.-J. Gabius and S. Gabius. Chapman & Hall (Weinheim), 1st edn, 1997, pp. 245–259.
26. S. Marionneau, A. Cailleau-Thomas, J. Rocher, B. Le Moullac-Vaidye, N. Ruvoën, M. Clément and J. Le Pendu. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie.,* (2001), **83**, 565-573.
27. a) S. Hakomori. Antigen structure and genetic basis of histo-blood groups A, B and O: their changes associated with human cancer. *Biochim. Biophys. Acta,* (1999), **1473**, 247–266. b) A. Cailleau-Thomas, P. Coullin, J. J. Candelier, L. Balanzino, B. Mennesson, R. Oriol and R. Mollicone. FUT4 and FUT9 genes are expressed early in human embryogenesis. *Glycobiology*, (2000), **10**, 789–802. c) P. Bengston, C. Börjeson, A. Lundblad, G. Larson and P. Pahlsson. A novel mutation in the human FUT7 gene. *Glycoconj. J.,* (1999), **16**, S48.
28. S. Weissbach, F. Flegge and T. Peters. Substrate binding drives active-site closing of human blood group B galactosyltransferase as revealed by hot-spot labeling and NMR spectroscopy experiments. *ChemBioChem,* (2018), **19**, 970–978.

29. C. Breton, L. Šnajdrová, C. Jeanneau, J. Koča and A. Imberty. Structures and mechanisms of glycosyltransferases. *Glycobiology,* (2006), **16**, 29R-37R.
30. K. Landsteiner. Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe". *Centralblatt f. Bakteriologie, Parasitenkunde u. Infektionskrankheiten,* (1900), **27**, 357–362.
31. F. Halary, A. Amara, H. Lortat-Jacob, M. Messerle, T. Delaunay, C. Houlès, F. Fieschi, F. Arenzana-Seisdedos, J. F. Moreau and J. Déchanet-Merville. Human cytomegalovirus binding to DC-SIGN is required for dendritic cell infection and target cell trans-infection. *Immunity,* (2002), **17**, 653-664.
32. F. Probert, S. B.-M. Whittaker, M. Crispin, D. A. Mitchell and A. M. Dixon. Solution NMR analyses of the C-type carbohydrate recognition domain of DC-SIGNR protein reveal different binding modes for HIV-derived oligosaccharides and smaller glycan fragments. *J. Biol. Chem.,* (2013), **288**, 22745-22757.
33. N. S. Stambach and M. E. Taylor. Characterization of carbohydrate recognition by Langerin, a C-type lectin of Langerhans cells. *Glycobiology,* (2003), **13**, 401-410.
34. G. Tabarani, M. Thépaut, D. Stroebel, C. Ebel, C. Vivès, P. Vachette, D. Durand and F. Fieschi. DC-SIGN neck domain is a pH-sensor controlling oligomerization: SAXS and hydrodynamic studies of extracellular domain. *J. Biol. Chem.,* (2009), **284**, 21229-21240.
35. H. Feinberg, Y. Guo, D. A. Mitchell, K. Drickamer and W. I. Weis. Extended neck regions stabilize tetramers of the receptors DC-SIGN and DC-SIGNR. *J. Biol. Chem.,* (2005), **280**, 1327-1335.
36. D. A. Mitchell, A. J. Fadden and K. Drickamer. A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR. Subunit organization and binding to multivalent ligands. *J. Biol. Chem.,* (2001), **276**, 28939-28945.
37. J. J. Reina, S. Sattin, D. Invernizzi, S. Mari, L. Martinez-Prats, G. Tabarani, F. Fieschi, R. Delgado, P. M. Nieto, J. Rojo and A. Bernardi. 1,2-mannobioside mimic: synthesis, DC-SIGN interaction by NMR and docking, and antiviral activity. *ChemMedChem,* (2007), **2**, 1030–1036.
38. K. Pederson, D. A. Mitchell, and J. H. Prestegard. Structural characterization of the DC-SIGN−LewisX complex. *Biochemistry,* (2014), **53**, 5700−5709.
39. a) C. J. Roberts. Non-native protein aggregation kinetics. *Biotechnol. Bioeng.,* (2007), **98**, 927-938. b) V. Vagenende, M. G. Yap and B. L. Trout. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry,* (2009), **48**, 11084-11096. c) T. Sharpe. Preventing protein aggregation. (*Biozentrum Biophysics Facility*), University of Basel. https://www.biozentrum.unibas.ch/fileadmin/redaktion/05_Facilities/01_Technology_Platforms/BF/Protocols/Preventing_Protein_Aggregation.pdf
40. Y. Zhang and P. S. Cremer. Interactions between macromolecules and ions: The Hofmeister series. *Curr. Opin. Chem. Biol.,* (2006), **10**, 658-663.

41. A. Holla and A. Skerra. Comparative analysis reveals selective recognition of glycans by the dendritic cell receptors DC-SIGN and Langerin. *Protein Eng. Des. Sel.,* (2011), **24**, 659-669.
42. P. Groves, M.O. Rasmussen, M.D. Molero, E. Samain, F. J. Cañada, H. Driguez and J. Jiménez-Barbero. Diffusion ordered spectroscopy as a complement to size exclusion chromatography in oligosaccharide analysis. *Glycobiology,* (2004), **14**, 451-456.
43. P. Groves, M. Palczewska, M. D. Molero, G. Batta. F. J. Cañada and J. Jiménez-Barbero. Protein molecular weight standards can compensate systematic errors in diffusion-ordered spectroscopy. *Anal. Biochem.,* (2004), **331**, 395-397.
44. C. Fernández and G. Wider. NMR spectroscopy of large biological macromolecules in solution. In *Advances techniques in biophysics*, eds. J. L. R. Arrondo and A. Alonso. Springer-Verlag (Berlín), 1st edn, 2006, **ch. 5**.
45. http://www.bmrb.wisc.edu. BMRB entry 19931: DC-SIGN. Authors: K. Pederson; D. A. Mitchell and J. H. Prestegard.
46. M. Ikura, L. E. Kay and A. Bax. A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry,* (1990), **29**, 4659-4667.
47. B. T. Farmer, R. A. Venters, L. D. Spicer, M. G. Wittekind and L. Müller. A refocused and optimized HNCA: increased sensitivity and resolution in large macromolecules. *J. Biomol. NMR,* (1992), **2**, 195-202.
48. K. Pervushin, R. Riek, G. Wider and K. Wüthrich. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U S A,* (1997), **94**, 12366-12371.
49. http://www.bmrb.wisc.edu.. BMRB entry 27854: Human DC-SIGN Carbohydrate Recognition Domain. Authors: P. Valverde, S. Delgado, J. D. Martinez, J. Jimenez-Barbero and A. Ardá.
50. D. Sharma and K. Rajarathnam. 13C NMR chemical shifts can predict disulfide bond formation. *J. Biomol. NMR,* (2000), **18**, 165-171.
51. M. V. Berjanskii and D. S. Wishart. Application of the random coil index to studying protein flexibility. *J. Biomol. NMR,* (2008), **40**, 31-48.
52. Y. Shen and A. Bax. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR,* (2013), **56**, 227-241.
53. M. P. Williamson. Using chemical shift perturbation to characterize ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.,* (2013), **73**, 1-16.
54. B. Meyer and T. Peters. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew. Chem. Int. Ed. Engl.,* (2003), **42**, 864-890.
55. J. Hanske, R. Wawrzinek, A. Geissner, E. C. Wamhoff, K. Sellrie, H. Schmidt, P. H. Seeberger and C. Rademacher. Calcium-Independent activation of an allosteric

network in Langerin by heparin oligosaccharides. *Chembiochem,* (2017), **18**, 1183-1187.

56. J. Aretz, H. Baukmann, E. Shanina, J. Hanske, R. Wawrzinek, V. A. Zapol`skii, P. H. Seeberger, D. E. Kaufmann and C. Rademacher. Identification of multiple druggable secondary sites by fragment screening against DC-SIGN. *Angew. Chem. Int. Ed. Engl.,* (2017), **56**, 7292-7296.
57. G. Timpano, G. Tabarani, M. Anderluh, D. Invernizzi, F. Vasile, D. Potenza, P. M. Nieto, J. Rojo, F. Fieschi and A. Bernardi. Synthesis of novel DC-SIGN ligands with an α-fucosylamide anchor. *ChemBioChem,* (2008), **9**, 1921–1930.
58. R. A. Williamson, M. D. Carr, T. A. Frenkiel, J. Feeney and R. B. Freedman. Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation. *Biochemistry,* (1997), **36**, 13882-13889.
59. Y. Guo, H. Feinberg, E. Conroy, D. A. Mitchell, R. Alvarez, O. Blixt, M. E. Taylor, W. I. Weis and K. Drickamer. Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat. Struct. Mol. Biol.,* (2004), **11**, 591–598.
60. H. Feinberg, R. Castelli, K. Drickamer, P. H. Seeberger and W. I. Weis. Multiple modes of binding enhance the affinity of DC-SIGN for high mannose N-linked glycans found on viral glycoproteins. *J. Biol. Chem.,* (2007), **282**, 4202–4209.
61. J. Hanske, S. Aleksić, M. Ballaschk, M. Jurk, E. Shanina, M. Beerbaum, P. Schmieder, B. G. Keller and C. Rademacher. Intradomain Allosteric Network Modulates Calcium Affinity of the C-Type Lectin Receptor Langerin. *J. Am. Chem. Soc.,* (2016), **138**, 12176-12186.
62. L. Fielding. NMR methods for the determination of protein-ligand dissociation constants. *Curr. Top. Med. Chem.,* (2003), **3**, 39-53.
63. J. Morrison, J. C. Yang, M. Stewart and D. Neuhaus. Solution NMR study of the interaction between NTF2 and nucleoporin FxFG repeats. *J. Mol. Biol.,* (2003), **333**, 587–603.
64. M. Reibarkh, T. J. Malia and G. Wagner. NMR distinction of single- and multiple-mode binding of small-molecule protein ligands. *J. Am. Chem. Soc.*, (2006), **128**, 2160-2161.
65. J. Feeney, J. G. Batchelor, J. P. Albrand and G. C. K. Roberts. The effects of intermediate exchange processes on the estimation of equilibrium constants by NMR. *J. Magn. Reson. (1969),* (1979), **33**, 519-529.
66. S. R. Haseley, P. Talaga, J. P. Kamerling and J. F. Vliegenthart. Characterization of the carbohydrate binding specificity and kinetic parameters of lectins by using surface plasmon resonance. *Anal. Biochem.,* (1999), **274**, 203-210.
67. J. D. Milton, D. G. Fernig and J. M. Rhodes. Use of a biosensor to determine the binding kinetics of five lectins for Galactosyl-N-acetylgalactosamine. *Glycoconj. J.,* (2001), **18**, 565-569.
68. M. Scharenberg, X. Jiang, L. Pang, G. Navarra, S. Rabbani, F. Binder, O. Schwardt and B. Ernst. Kinetic properties of carbohydrate-lectin interactions: FimH antagonists. *ChemMedChem,* (2014), **9**, 78-83.

69. A. R. Fersht. *Structure and Mechanism in Protein Science*. W. H. Freeman (New York), 1st edn, 1999.
70. H.-C. Siebert, S. André, S. Y. Lu, M. Frank, H. Kaltner, J. A. van Kuik, E. Y. Korchagina, N. Bovin, E. Tajkhorshid, R. Kaptein, J. F. Vliegenthart, C. W. von der Lieth, J. Jiménez-Barbero, J. Kopitz and H. J. Gabius. Unique conformer selection of human growth-regulatory lectin galectin-1 for ganglioside GM1 versus bacterial toxins. *Biochemistry,* (2003), **42**, 14762–14773.
71. H. Wegener, A. Mallagaray, T. Schöne, T. Peters, J. Lockhauserbäumer, H. Yan, C. Uetrecht, G. S. Hansman and S. Taube. Human norovirus GII.4(MI001) P dimer binds fucosylated and sialylated carbohydrates. *Glycobiology,* (2017), **27**, 1027–1037.
72. A. Mallagaray, C. Rademacher, F. Parra, G. Hansman and T. Peters. Saturation transfer difference nuclear magnetic resonance titrations reveal complex multistep binding of L-fucose to norovirus particles. *Glycobiology,* (2017), **27**, 80–86.
73. J. D. Martínez, P. Valverde, S. Delgado, C. Romanò, B. Linclau, N. C. Reichardt, S. Oscarson, A. Ardá, J. Jiménez-Barbero and F. J. Cañada. Unraveling Sugar Binding Modes to DC-SIGN by Employing Fluorinated Carbohydrates. *Molecules,* (2019), **24**, E2337.
74. H. Feinberg, M. E. Taylor, N. Razi, R. McBride, Y. A. Knirel, S. A. Graham, K. Drickamer and W. I. Weis. Structural basis for langerin recognition of diverse pathogen and mammalian glycans through a single binding site. *J. Mol. Biol.,* (2011), **405**, 1027-1039.
75. K. K. Ng, K. Drickamer and W. I. Weis. Structural analysis of monosaccharide recognition by rat liver mannose-binding protein. *J. Biol. Chem.,* (1996), **271**, 663-674.
76. a) P. J. Kraulis. ANSIG: A program for the assignment of protein 1H 2D NMR spectra by interactive graphics. *J. Magn. Reson. (1969),* (1989), **84**, 627-663. b) S. P. Skinner, B. T. Goult, R. H. Fogh, W. Boucher, T. J. Stevens, E. D. Laue and G. W. Vuister. Structure calculation, refinement and validation using CcpNmr Analysis. *Acta Crystallogr. Sect. D: Biol. Crystallogr.,* (2015), **71**, 154-161.
77. O. Zerbe and S. Jurt. *Applied NMR Spectroscopy for Chemists and Life Scentists*. Wiley-VCH Verlag GmbH & Co. KGaA (Weinheim), 1st edn, 2014.
78. C. Guzzi, J. C. Muñoz-García, P. M. Enriquez-Navas, J. Rojo, J. Angulo and P. M. Nieto. NMR studies on carbohydrate interactions with DC-SIGN towards a quantitative STD analysis. *Pure Appl. Chem.,* (2013), **85**, 1771–1787.
79. C. Guzzi, P. Alfarano, I. Sutkeviciute, S. Sattin, R. Ribeiro-Viana, F. Fieschi, A. Bernardi, J. Weiser, J. Rojo, J. Angulo and P. M. Nieto. Detection and quantitative analysis of two independent binding modes of a small ligand responsible for DC-SIGN clustering. *Org. Biomol. Chem.,* (2016), **14**, 335-344.
80. E. van Liempt, A. Imberty, C. M. C. Bank, S. J. van Vliet, Y. van Kooyk, T. B. Geijtenbeek and I. van Die. Molecular basis of the differences in binding properties of the highly related C-type lectins DC-SIGN and L-SIGN to Lewis X

trisaccharide and *Schistosoma mansoni* egg antigens. *J. Biol. Chem.,* (2004), **279**, 33161-33167.

81. S. P. Arepalli, C. P. Glaudemans, G. D. Daves, P. Kovac and A. Bax. Identification of protein-mediated indirect NOE effects in a disaccharide-Fab' complex by transferred ROESY. *J. Magn. Reson. Ser. B,* (1995), **106**, 195-198.
82. K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.,* (2008), **29**, 622-655.
83. B. L. Foley, M. B. Tessier and R. J. Woods. Carbohydrate force fields. *Wiley Interdiscip. Rev. Comput. Mol. Sci.,* (2012), **2**, 652-697.
84. R. Boelens, T. M. G. Koning and R. Kaptein. Determination of biomolecular structures from proton-proton NOE's using a relaxation matrix approach. *J. Mol. Struct.,* (1988), **173**, 299-311.
85. R. Boelens, T. M. G. Koning, G. A. der Marel, J. H. van Boom and R. Kaptein. Iterative procedure for structure determination from proton-proton NOEs using a full relaxation matrix approach. Application to a DNA octamer. *J. Magn. Reson. (1969),* (1989), **82**, 290-308.
86. M. H. Levitt. *Spin dynamics: basics of nuclear magnetic resonance*. John Wiley & Sons (Chichester), 2nd edn, 2008.
87. S. J. Angyal. The composition and conformation of sugars in solution. *Angew. Chem. Int. Ed.,* (1969), **8**, 157-166.
88. B. R. Seavey, E. A. Farr, W. M. Westler and J. L. Markley. A relational database for sequence-specific protein NMR data. *J. Biomol. NMR,* (1991), **1**, 217-236.
89. A. Bundi and K. Wüthrich. ^{1}H-NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers,* (1979), **18**, 285-297.
90. D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges and B. D. Sykes. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR,* (1995), **5,** 67-81.
91. C. R. Cantor and P. R. Schimmel. *Biophysical Chemistry Part II: Techniques for the Study of Biological Structure and Function*. Freeman and Co. (San Francisco), 1980, p. 461.
92. Y. Yuan, D. W. Bleile, X. Wen, D. A. Sanders, K. Itoh, H. W. Liu and B. M. Pinto. Investigation of binding of UDP-Galf and UDP-[3-F]Galf to UDP-galactopyranose mutase by STD-NMR spectroscopy, molecular dynamics, and CORCEMA-ST calculations. *J. Am. Chem. Soc.,* (2008), **130**, 3157–3168.
93. N. R. Krishna and V. Jayalakshmi. Complete Relaxation and Conformational Exchange Matrix Analysis of STD-NMR spectra of ligand receptor complexes. *Prog. Nucl. Magn. Reson. Spectrosc.,* (2006), **49**, 1-25.
94. Y. Imahori, T. Ido, K. Ishiwata, T. Takahashi, K. Kawashima, K. Yanai, Y. Miura, M. Monma, S. Watanuki and A. Ujiie. 2-deoxy-2-[18F]-fluoro-L-fucose, a potential agent for regional fucose utilization studies associated with glycoprotein synthesis. *CYRIC Annual Report,* (1984), 119-131.

95. C. G. Butchard and P. W. Kent. 2-deoxy-2-fluoro-L-fucose: synthesis and structure. *Tetrahedron,* (1979), **35**, 2551-2554.
96. T. Haradahira, M. Maeda, Y. Yano and M. Kojima. New synthesis of 2-deoxy-2-fluoro-D-galactose. *Chem. Pharm. Bull.,* (1984), **32**, 3317-3319.
97. K. Koch and R. J. Chambers. An improved synthesis of 4-deoxy-4-fluoro-d-galactopyranosyl derivatives. *Carbohydr. Res.,* (1993), **241**, 295–299.
98. J. R. Rasmussen, S. R. Tafuri and S. T. Smale. Synthesis of 3-deoxy-3-fluoro-d-mannose and 4-deoxy-4- fluoro-d-mannose. *Carbohydr. Res.,* (1983), **116**, 21–29.
99. C. Dalvit. Theoretical analysis of the competition ligand-based NMR experiments and selected applications to fragment screening and binding constant measurements. *Concepts Magn. Reson.,* (2008), **32A**, 341-372.
100. H. Y. Carr and E. M. Purcell. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.,* (1954), **94**, 630-638.
101. S. Meiboom and D. Gill. Modified Spin-Echo method for measuring nuclear relaxation times. *Rev. Sci. Instrum.,* (1958), **29**, 688-691.

CHAPTER 5: DC-SIGN RECOGNITION OF FUCOSYLATED ANTIGENS FROM S. MANSONI

This work has been performed in collaboration with Prof. G. J. Boons at the University of Utrecht.

Schistosomes are parasitic helminths well known for their ability to manipulate the host's immune system and establish chronic infections, and are responsible for 280,000 deaths annually and over 250 million infected people worldwide [1-2]. As many other pathogens, schistosome parasites present a wide plethora of complex N-glycans and O-glycans coating their cell surfaces, most of them fucosylated as those from mammals [3]. From a biological perspective, DC-SIGN has been already identified as a relevant receptor targeted by *S. mansoni* glycans during schistosomiasis [4]. However, the glycan profile displayed by these organisms is remarkably extensive and has been shown to significantly vary among male and female individuals and also along the different stages of the parasite development [5]. Moreover, the immune-modulatory properties of the soluble egg antigen (SEA) of *S. mansoni* have been mainly examined by using glycoprotein mixtures and as a result, the potential effects of individual glycans in immunity shaping have been difficult to study [6].

Besides LacNAc (LN) and Le^X, the SEA of *S. mansoni* contains a series of sugar antigens that are less common or inexistent in mammals, and by contrast, they completely lack sialic acid as terminal modification. All these distinct antigens share the common disaccharide scaffold LacDiNAc (LDN), which is barely present in vertebrates [7]. This simple motif can be then likewise modified by fucosyltransferases, giving rise to several mono-Fuc- and di-Fuc-containing structures, including LDN-F, F-LDN, LDN-DF, F-LDN-F and DF-LDN (Figure 5.1A) [3,8]. LDNF and Le^X are actually two representative epitopes well documented on glycoproteins and glycolipids from the three major schistosome species [9]. However, while the latter is easily found in both mammals and parasites, the presence of the former in mammal cells is very restricted (only found in two glycoproteins till now) [10]. Regarding the polyfucosylated structures, they have never been described in any other parasitic or mammalian host species, but are known to trigger potent antibody responses in humans and primates [11].

Interestingly, two of the aforementioned glycans, LDNF and LDN-DF, have been already investigated as potential ligands of DC-SIGN. Early results evidenced that LDNF is as good DC-SIGN binder as Le^X, whereas LDN-DF is strikingly poorly recognized [12]. Taking into account that the structures of

LDNF and Le^X are almost equivalent, it could be hypothesized that the recognition takes place in a similar way and that the presence of a N-acetyl group in LDNF would not affect the recognition event. However, it remains surprising that a di-fucosylated antigen such as LDN-DF can escape from detection by DC-SIGN. No structural studies had explained these observations yet.

In this scenario, the present chapter has been focused on shedding light into this particular question. Thus, the molecular recognition processes of DC-SIGN towards these two glycan motifs has been individually analyzed by NMR, using the same approach followed in the previous chapter (Figure 5.1B). The results should be useful to establish the molecular basis on how *S. mansoni* parasites use these glycans to trigger or avoid DC-SIGN-mediated responses on its benefit.

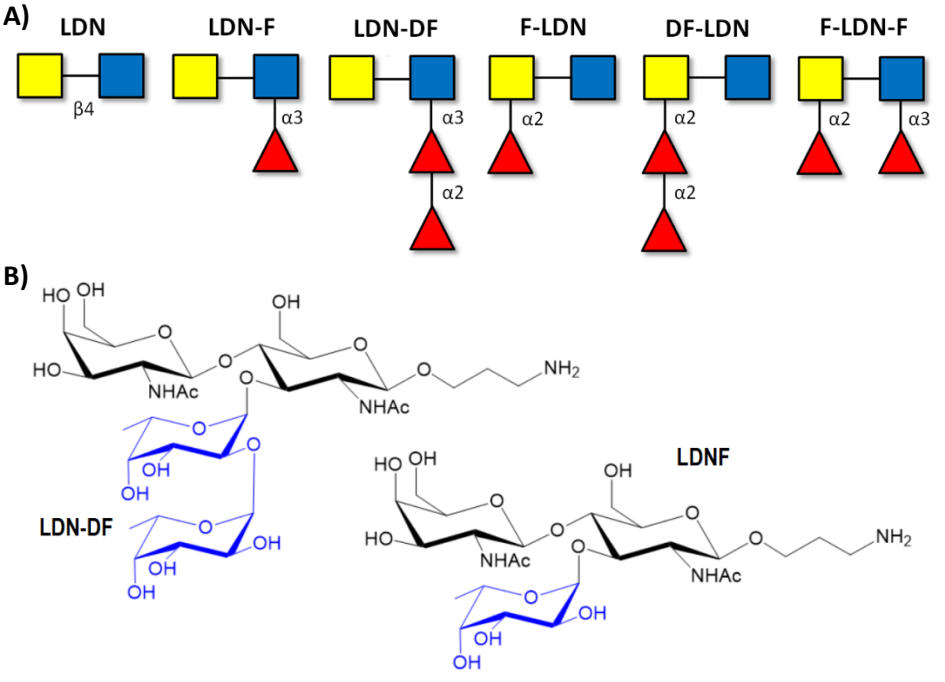


Figure 5.1. *A) Common glycan structures found in Schistosoma mansoni egg proteins. B) Structures of the two antigens employed for molecular recognition studies.*

5.1. HSQC-based titrations

The recognition of the two ligands was firstly evaluated from the receptor´s point of view, through the ^{15}N-HSQC spectra. The titrations were performed with the TROSY pulse sequence and adding relatively high amounts of ligand,

ranging from tens to above one hundred equivalents, thereby maximizing the resolution and enabling the detection of noticeable perturbations to build the corresponding binding isotherms (310 K).

In particular, seven titration points were recorded for each ligand, and two completely different scenarios were found in each case. For the LDNF trisaccharide, protein saturation was almost reached around 70-80 eq of ligand, hence optimal CSPs could be perfectly calculated at 100 eq (Figure 5.2). On the contrary, the LDN-DF tetrasaccharide yielded very weak perturbations at the same relative concentration. Given this particular scenario, more titrations points were acquired for the LDN-DF by increasing its relative amount up to ca. 300 eq. However, even at this concentration a fully saturated binding curve could not be reached (Figure 5.5). These observations fittingly assessed the reported extremely low affinity of DC-SIGN for this antigen [12].

In both cases, the set of CSPs above the standard threshold (mean + stdev.) were all located around the calcium site 2 (Figure 5.3), thus demonstrating the specificity of such interactions. Moreover, both profiles, especially that of the LDNF, were significantly similar to those of the blood groups (Figure 5.2), with minor differences. For the LDNF antigen, most of its CSPs actually resembled those from the B antigen, except for some interesting discrepancies. As highlights, perturbations measured for D367 and K368 resulted fairly larger than those for both blood groups. In fact, they resembled those measured for the Fuc monosaccharide. In addition, the pattern of long-range CSPs slightly changed as well: intriguingly, T314 was barely affected upon ligand binding, whereas it has been reported so far to be a common shifted residue by both Fuc- and Man-containing ligands. The same fact has been observed for G361, at the border of the binding site. Alternatively, the R309 residue, which is located at the top of the α_2-helix (Figure 5.3), was definitely affected by the ligand in spite of being placed far away from the binding site. In contrast, for LDN-DF, no realistic observations could be derived from the CSP analysis in the same conditions (100 eq), since many residues uniformly distributed throughout the entire protein construct suffered from notable perturbations above or close to the theoretical threshold, due to the low data dispersion. Fortunately, some few residues eventually stood out upon adding ca. 200 eq more, thereby proving some specificity of the interaction, as mentioned above. In particular, only eight residues could be considered as affected by specific binding (Figure 5.3).

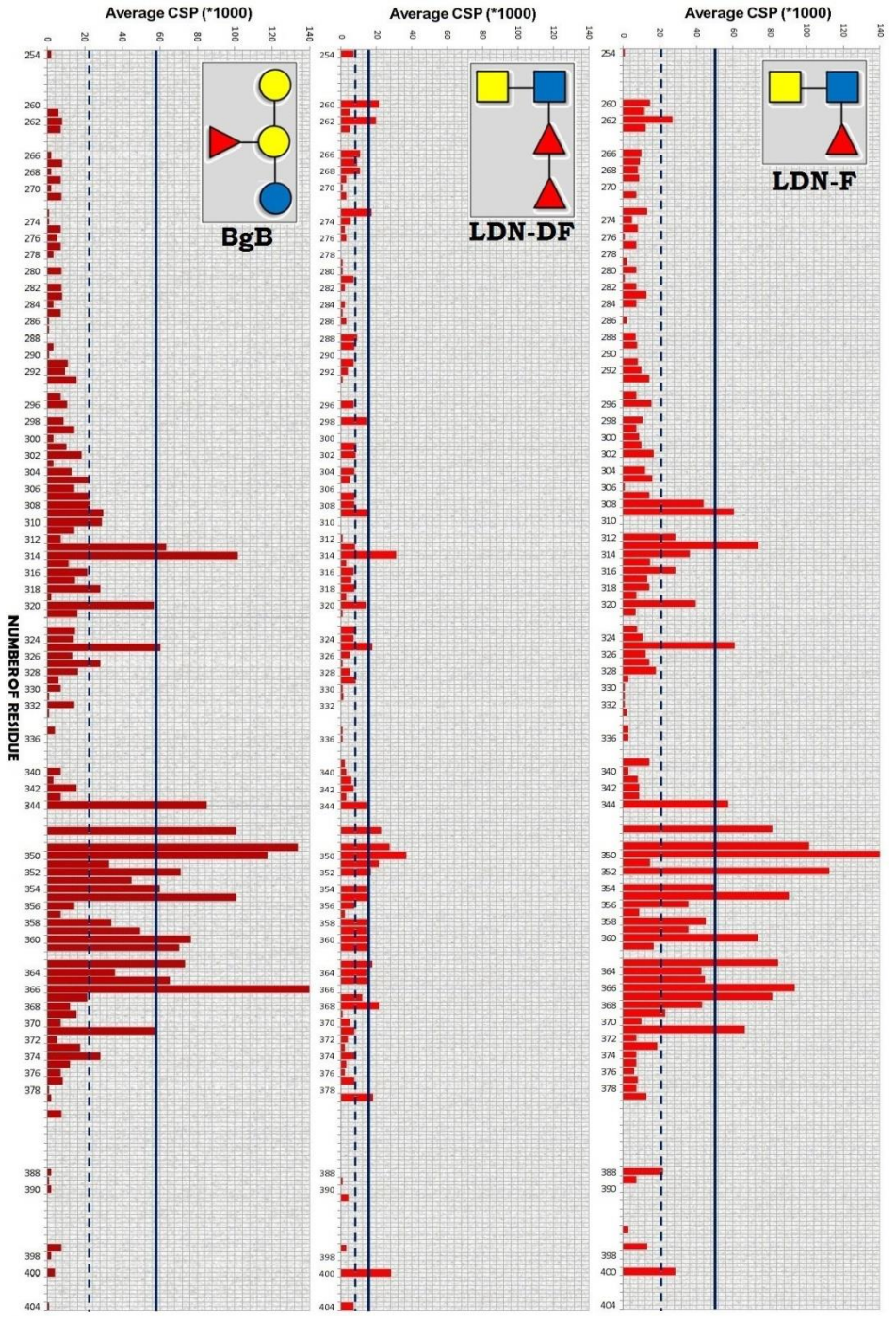
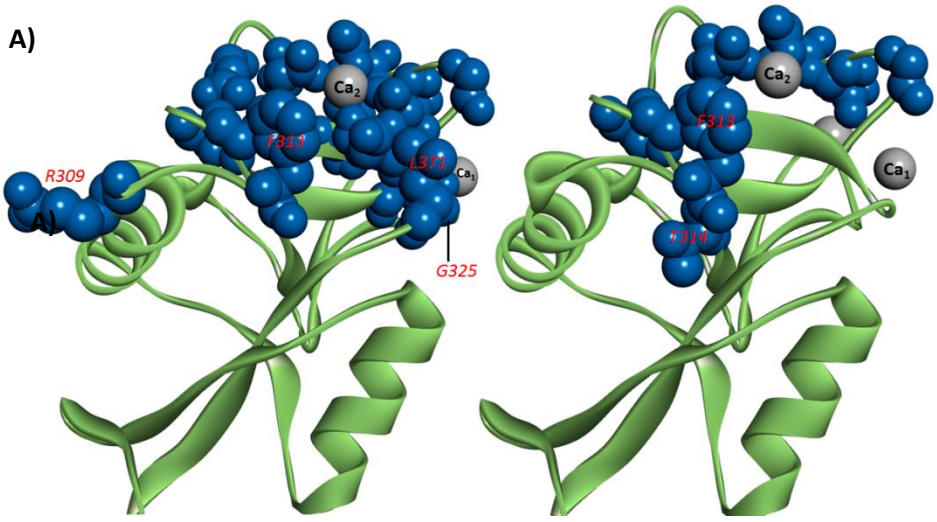


Figure 5.2. *Average chemical shift perturbations calculated through Eq. 8 using α = 0.14 (See Methods). For LDNF and LDN-DF, CSPs are estimated in the presence of 100 eq of ligand. For BgB, values are derived at 140 eq. In all cases, the working temperature was 310 K.*

A)



B)

	Mean (μ)	Stdev. (σ)	Cutoff (μ+σ)	CSPs>Cutoff
Blood Group B (140eq)	21.7	35.4	57.2	**16**
LDNF (100eq)	20.1	27.2	47.4	**15**
LDN-DF (100eq)	7.0	7.6	14.6	**8***

**Checked at 295 eq.*

Figure 5.3. *A) CRD structure showing those amino acids which yielded the strongest CSPs (Left: LDNF. Right: LDN-DF). Red labels indicate those perturbed amino acids located at relatively remote positions from the binding site. Calcium ions are represented as white spheres. B) Statistical analysis of the collected CSPs for the two fucosylated ligands studied herein as compared with those previously shown for the blood group B (values displayed as ppm*1000).*

Regarding the binding kinetics, the crosspeak evolution observed for both antigens strongly suggested the existence of an intermediate exchange regime on the chemical shift scale. Thus, in both cases, some of the most perturbed peaks were observed to deeply broaden at the intermediate points of the titration. Then, at a certain point, these peaks were progressively becoming sharper again and the signal intensity was partially recovered (Figure 5.4). Interestingly, as observed for the A antigen, D366 was highly affected by these broadening phenomena and quickly disappeared in the presence of few equivalents of ligand. Its evolution was likewise harder to track since it only re-appeared as a much weakened peak when protein saturation had been almost reached. Moreover, for LDNF, a similar effect was also noticed for F313. Overall, the intermediate exchange could be readily observed for the LDNF antigen, as coalescence effects were more acute, giving rise to the complete loss of several amino acid crosspeaks along the titration (ca. 7-8 residues).

Instead, the LDN-DF displayed smoother peak broadening effects, although still detectable. Obviously, peak recovery could not be observed in all cases, as complete saturation had not been achieved yet (for instance, F313 & N349, Figure 5.4).

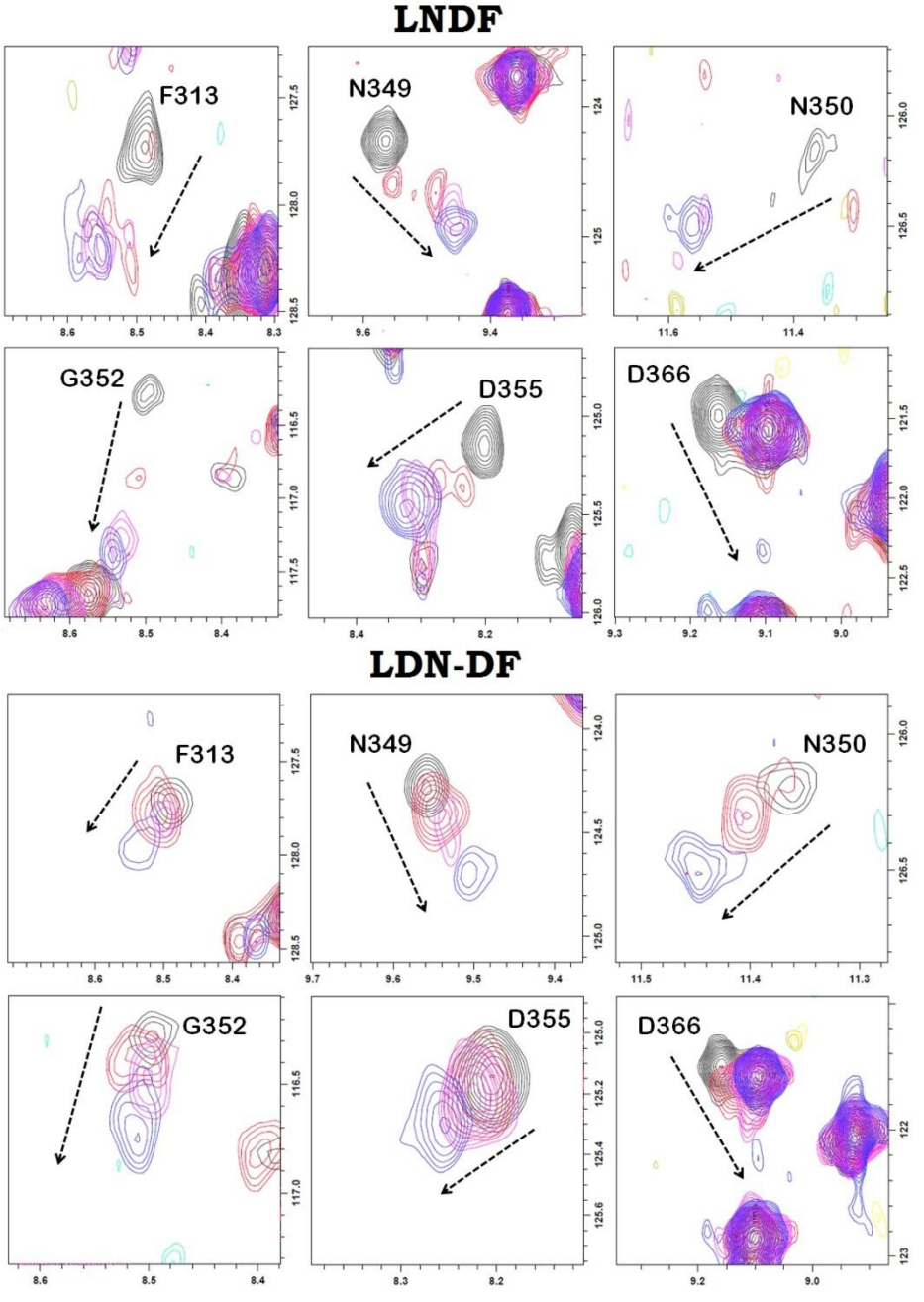


Figure 5.4. *Perturbations monitored for six representative amino acids. The four titration points represented in black, red, pink and blue correspond, respectively, to 0, 20, 60 and 100 eq of LDNF and to 0, 50, 100 and 295 eq of LDN-DF.*

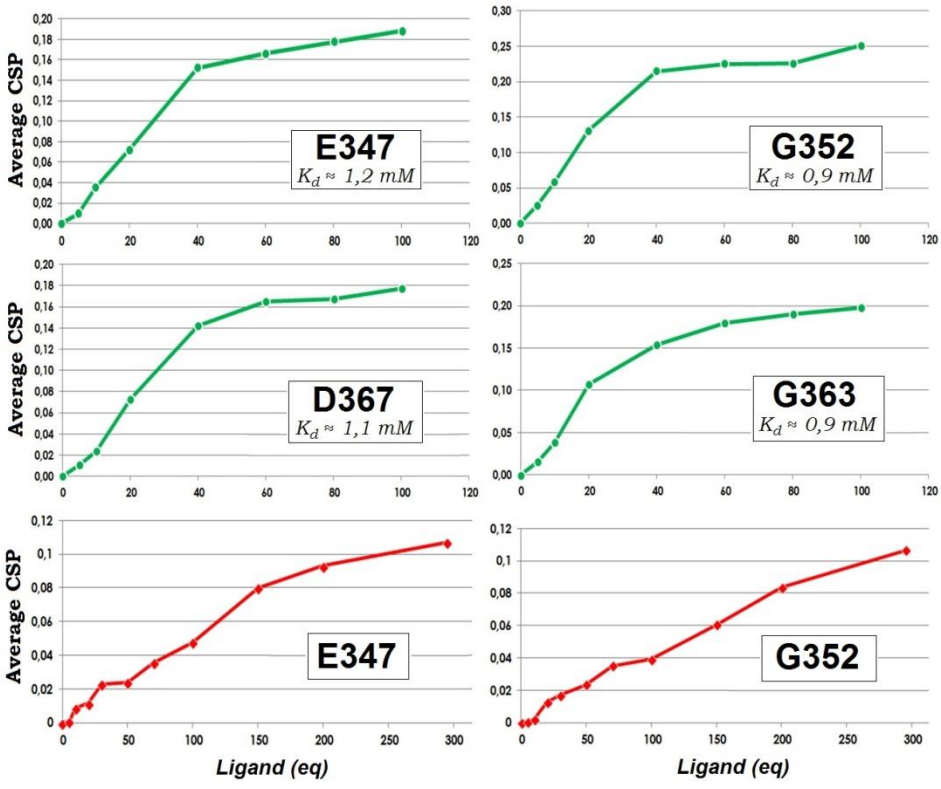


Figure 5.5. *Binding isotherms built for several residues notably shifted (310 K). The green lines correspond to the LDNF antigen and the red lines to the LDN-DF ligand. Binding affinities were estimated through the equation* $y = A(B+x-[(B+x)^2-4x]^{0.5})$, *with* $A = \Delta\delta_{max}/2[P_0]$ *and* $B = [P_0] + K_d$.

In any case, the peak trajectories were all linear and therefore, indicative of a simple binding event (stoichiometry 1:1) at the primary binding site. Analogously, the binding isotherms obtained for the LDNF trisaccharide showed a normal growing profile, without intermediate steps. As illustrated in the previous chapter, this behavior discards the existence of multiple binding events either at the same binding site or at different protein sites [13-14]. However, it is worth mentioning a particular difference between the build-up curves derived for LDNF and for the blood group antigens. In the latter case, although the intermediate exchange could be verified by the estimated off rate kinetic constants and the coalescence phenomena, the profiles described a typical exponential growing, which actually better fits with a fast exchange behavior. As reported [15], when the binding kinetics of a ligand becomes slower, its corresponding binding isotherms progressively tend to adopt a sigmoidal profile. Indeed, this was the case for LDNF (Figure 5.5): the most perturbed residues clearly displayed sigmoidal profiles at low protein-ligand

ratios. With the global CSP data, the curve fitting was carried out by using Eq. 10 described in *Methods* and the dissociation constant was estimated as an average of the individual values from each case. To highlight, the intense peak broadening precluded the use of many of the most affected amino acids which still yielded measurable but very irregular peaks at the intermediate titration points, whereby the peak centers could not be accurately determined. Eventually, four amino acids were suitably employed to give a feasible estimation of the dissociation constant, with a value around 1.0 ± 0.5 mM. This value is in the same range that the affinities found for the blood groups (*Chapter 4*) and for the Lewis X antigen [16]. Moreover, as can be visually checked in Figure 5.2, the absolute values of the CSPs measured for the LDNF at 100 eq were of the same order that the same values obtained for the B antigen at 140 eq. Therefore, a slightly higher affinity of the former (1.0 mM) in comparison with the latter (2.3 mM) was reasonably expected.

Regarding the LDN-DF tetrasaccharide, since the absolute CSP values measured for this antigen were extremely low, the binding isotherms presented an irregular profile due to the measurement inaccuracies, even using the most perturbed residues for the calculations (E347 and G352, Figure 5.5). In either case, as the CRD could not be properly saturated, the corresponding dissociation constant was not estimated. As a rough prediction, the curve fitting to the available data afforded a K_d value about 10 mM. Hence, the real affinity is expected to be above this value.

$$\mathbf{k_{ex}} = \frac{\boldsymbol{\pi} \cdot \boldsymbol{\Delta\omega}}{\sqrt{2}} = \boldsymbol{k_{on}[L] + k_{off}}$$

	N349			N350/G352		
	Δω Hz	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})	Δω Hz	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})
10eq BgA ***$K_d \approx 7.1$ mM***	88	**$2.3 \cdot 10^4$**	**$1.7 \cdot 10^2$**	94	**$2.5 \cdot 10^4$**	**$1.8 \cdot 10^2$**
10eq BgB ***$K_d \approx 2.3$ mM***	107	**$6.8 \cdot 10^4$**	**$1.6 \cdot 10^2$**	94	**$6.0 \cdot 10^4$**	**$1.4 \cdot 10^2$**
20eq LDNF ***$K_d \approx 1.0$ mM***	200	**$10.0 \cdot 10^4$**	**$1.0 \cdot 10^2$**	257	**$12.8 \cdot 10^4$**	**$1.3 \cdot 10^2$**

Figure 5.6. *Estimation of the binding kinetics for LDNF as compared with the data previously estimated for the blood groups. N350 could not be used in this particular case, and G352 was alternatively used. The given values correspond to the same measuring temperature, 310 K.*

Finally, the binding kinetics was investigated for LDNF by means of Eq. 7 (*Methods*), as shown in Figure 5.6. Using the calculated dissociation constant, the resulting k_{on} and k_{off} values suggested that the slightly better affinity of this ligand if compared with the blood groups is essentially contributed by a larger on-rate. In principle, these results would attribute the LDNF trisaccharide a better ability to enter the binding site. Of course, the off-rate had a value in accordance with that expected from a system within an intermediate exchange regime [17]. To clarify, the coalescence point in this case took place around 1 mM (slightly lower than for the blood group antigens), corresponding to 20 eq of ligand instead of 10 eq (Figure 5.6) due to a lower amount of CRD used during NMR acquisition (50 µM).

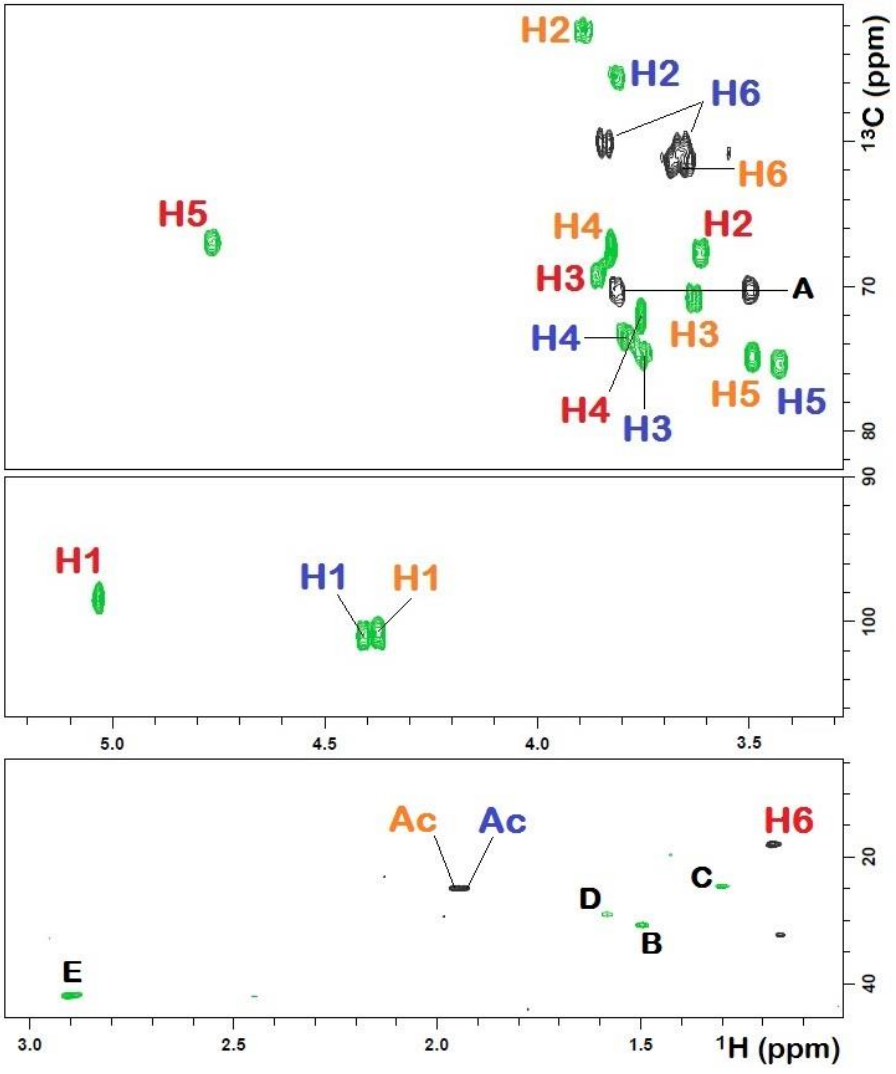
5.2. STD and tr-NOESY

To explore the binding pose of the ligands, more information was gathered from the ligand´s point of view. In essence, this information was derived from STD NMR and tr-NOESY experiments. Alike to the blood groups (See *Section 4.3*), the HSQC-based titrations have pointed out that the interaction of both ligands is basically contributed by the Fuc residue, given the strong similarities between the CSP patterns from the Fuc monosaccharide and its related oligosaccharides, including both the two blood group antigens and the two LDN-based antigens. However, as remainder, these observations do not necessarily imply that there will not be any secondary contacts established by the surrounding sugars. Certainly, despite the importance of such contacts, they are roughly or not detectable by ^{15}N-HSQC experiments, especially considering the weak nature of the interactions studied herein. Indeed, the slight differences found between fucosylated ligands could arise from these secondary non-covalent bonds, but in no case precise information can be directly inferred on the spatial arrangement of the oligosaccharide. Instead, these data could be jointly deduced from STD NMR and tr-NOESY results.

Firstly, the NMR spectra of both oligosaccharides were fully assigned through the same methodology employed for the blood group antigens. To note, for both structures the β-anomer of the reducing sugar (core GlcNAc) was fixed by a synthetic aglycone (n-aminopentyl chain), thus easing the identification of the anomeric protons (two β and one α). From this point, the spin system of each sugar residue was individually identified and assigned by means of their TOCSY patterns (Figure 5.7). The Fuc ring could be straightforwardly assigned due to both the characteristic chemical shift from the methyl group at C6 and the only α-proton located at the anomeric region. The other two patterns

were rather similar, as both displayed their characteristic H2 protons at higher field in the ^{13}C dimension, as a result of the N-acetylation. Luckily, the broad singlet shape commonly displayed by H4 in Gal pyranoses allowed to undoubtedly identifying the GalNAc moiety. In parallel, H3 from the central GlcNAc ring could be directly correlated with H1 Fuc in the 2D-HMBC, thereby confirming the proposed assignment. The remaining signals with opposite sign (CH_2) were assigned to the methylene groups from the aglycone chain. The assignment of the LDN-DF tetrasaccharide was fairly trivial, considering that only the chemical shifts from the analogous Fuc were affected by the presence of the other neighboring Fuc (the terminal one now).

A)



**Note: peak colors are inverted in the bottom window.*

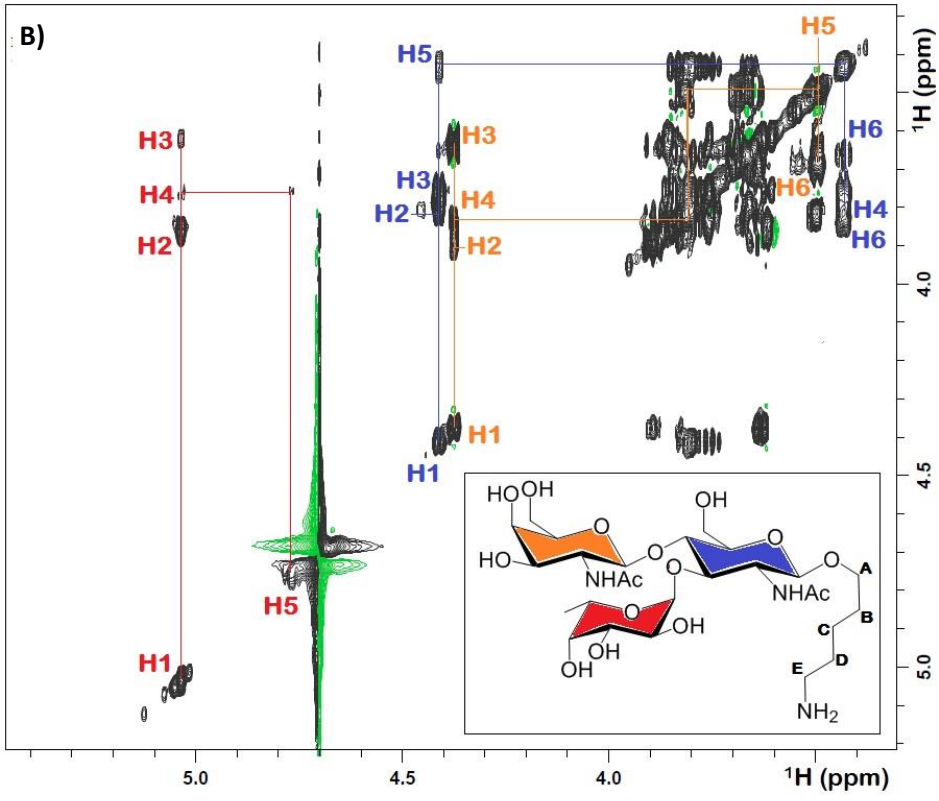


Figure 5.7. *A) ^{13}C-HSQC full assignment of the LDNF trisaccharide. Blue: β-GlcNAc; orange: β-GalNAc; red: α-Fuc. As noted, the aliphatic region containing Fuc H6, the N-acetyl groups and the aglycone methylenes is shown on the bottom. B) 2D-TOCSY assignment performed for the same antigen at 298 K. Each independent spin system is depicted in a different color and corresponds to a single sugar.*

Worth noting, the Fuc moiety α1-3-linked to the core GlcNAc presented in both cases an abnormally high chemical shift for H5 (4.7-4.8 ppm), ranging within the same values observed for anomeric protons (Figure 5.7). This is a particular feature of the Lewis-type antigens, which are known to adopt highly rigid structures in solution [18] with the non-reducing residue (often Gal or GalNAc) strongly stacking on top of the nearby Fuc. As a result of such an interaction, a "non-conventional" H-bond has been noticed to exist between H5 and the ring oxygen of the stacking sugar [19-20], which is responsible for the observed deshielding of the corresponding Fuc H5.

Considering that the N-acetyl group from the core GlcNAc has been reported to be key in the recognition of Lewis-type antigens [21], the two N-acetyl groups were also independently identified through a 2D-NOESY spectrum in H_2O, by means of correlating the corresponding H2 proton through the intermediate NH group.

Then, the STD NMR experiments were carried out. To enable a realistic comparison between the data collected for the blood groups and those discussed herein, the same experimental setup was put in practice. Namely, the spectra were acquired with a consensus saturation pulse length of 2 sec. applied within the aliphatic region of the protein (0.76 ppm). As already stated, these parameters were optimal to guarantee an efficient transference of saturation, specially arising from the close Val351 sidechain, which can provide intense enough STDs without severe biasing effects arising from the different evolution of the saturation profile that may occur at longer saturation times. In addition, the recording was performed with the same protein:ligand ratios (1:70), using the CRD. In regard to the working temperature, H5 Fuc (4.7-4.8 ppm) was the most problematic nucleus. Expectedly, any change in the working temperature did not permit to separate this signal from the residual water signal (~10 mM), and eventually, the same temperature used for the characterization was maintained (298 K), which allowed the clean measuring of at least the three anomeric protons. Also, no water suppression was applied, but instead, the corresponding STDD spectra were obtained by subtracting a reference STD of the sole protein in the same conditions.

A qualitative analysis of the STD effects was carried out by measuring the corresponding peak intensities. For LDNF, a clear set of signals could be distinguished, whereas for the LDN-DF only some few saturated nuclei were detectable. The overall results are summarized as relative STD values in Figure 5.8. Noteworthy, the absolute STD intensities were substantially higher for LDNF than those described for the blood groups: specifically, the highest STD percentages noticed in the former case reached 6-7% of saturation, in comparison with the 1-2% of maximum intensity yielded by the blood group antigens. Interestingly, these found values were within the same order than those previously reported for the close Le^X analogue (5-6%) [21]. In principle, a higher STD intensity is equivalent to a slower k_{off}, which often correlates with a better affinity (lower K_d) [22]. In this regard, these observations supported that the difference between the k_{off} values estimated for the blood groups and for the LDNF was meaningful (Figure 5.6), and hence, the k_{off} of LDNF is actually lower and contributes to its improved affinity.

Going into detail, some similarities and some discrepancies were noticed for the STD profile of the LDNF respect to that of the Le^X trisaccharide [21]. In fact, the most important difference concerned the STD effects found for the interacting Fuc pyranose. Undoubtedly, the most affected protons by the

transference of saturation were Fuc H1 and Fuc H2, but in this case, Fuc H1 clearly stood out as the most perturbed nucleus. As already commented (*Chapter 4*), the scenario described for both the blood groups and the Le^X antigen was just the opposite. Strikingly, Fuc H4 has been even reported [21] to yield a higher STD than Fuc H1 in Le^X, which result a weird finding taking into account the relative arrangement of Fuc respect to Val351 when the Ca^{2+} ion is targeted through hydroxyls OH3 and OH4 (Figure 5.9B). In either case, the values reported herein are in accordance with the involvement of both protons (H1/H2) in close contacts with the Val351 sidechain. Apart, the STDs from Fuc H3, H4 and the methyl group were significantly lower, as found for the blood groups.

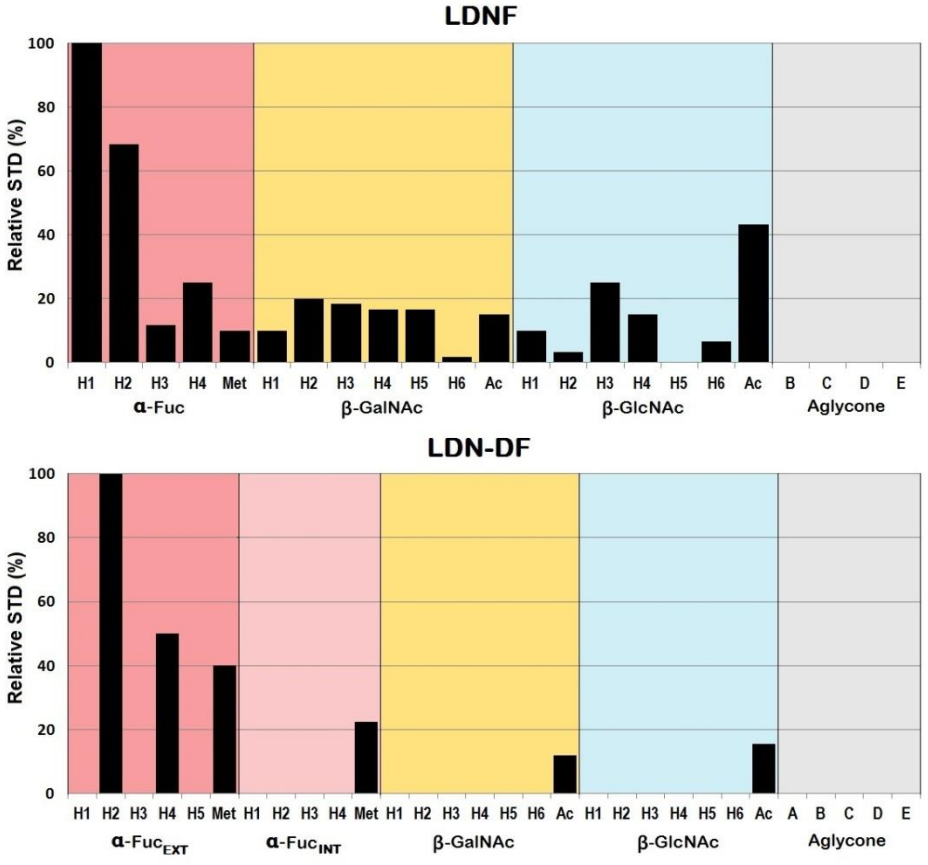
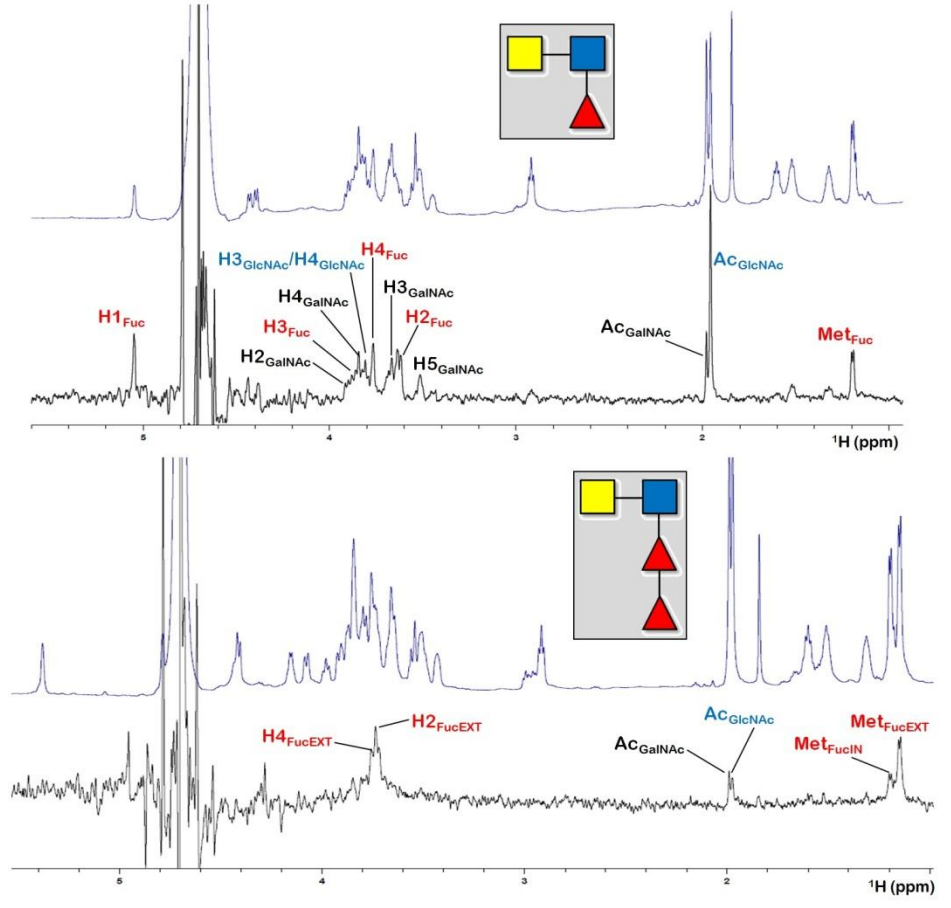


Figure 5.8. *Comparative analysis of the relative STD intensities obtained for both Lewis-type antigens in the presence of the DC-SIGN CRD. For the LDNF, Fuc H1 was established as the reference nucleus (100% rel.STD), and for LDN-DF, Fuc H2.*

Regarding the core GlcNAc, a clear consensus between the found STD and that reported for Le^X pointed out the closeness of the N-acetyl group to the Val351 residue, being actually the third most saturated position behind Fuc H1 and

H2. The remaining nuclei also received some saturation, although the relative percentages were notably lower as compared with that of the cited N-acetyl group. In the same line, a set of weak STD could be extracted from the non-reducing GalNAc peaks, namely H2, H3, H4, H5 and the N-acetyl group (Figure 5.9A). Even low, these STDs might be meaningful, as all of them ranged around 10-20% of intensity in relative terms, similar to those described for the terminal Gal in the blood group antigens. Thus, given that this Gal moiety contacted the protein surface in the blood group antigens, as demonstrated by tr-NOESY spectra and MD calculations, this possibility could not be discarded for LDNF. In fact, none of the three residues were completely devoid of significant STDs, suggesting that the full ligand structure was arranged such that the three sugars are contacting the protein surface. In addition, the resemblance between the STD profiles from LDNF and Le^X strongly supported the possibility that both trisaccharides share a very close, if not the same, binding pose.

A)



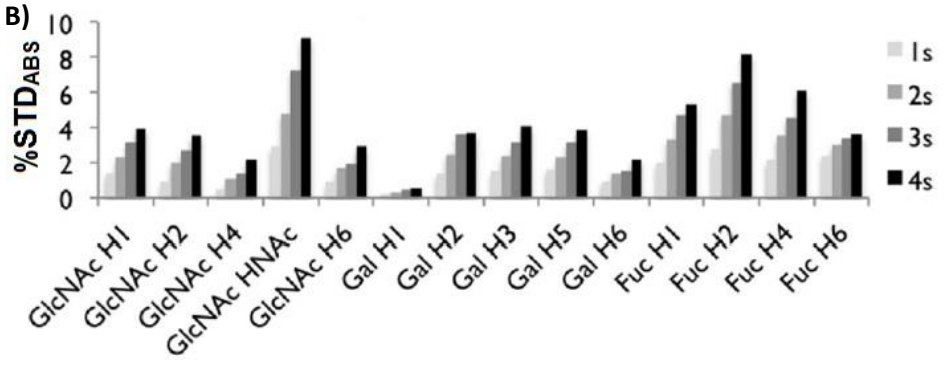


Figure 5.9. *A) Assignment of the most notorious STD effects detected for each antigen (Top: LDNF. Bottom: LDN-DF). Blue: β-GlcNAc; black: β-GalNAc; red: α-Fuc. B) Reported STD values for the Le^X antigen in complex with the DC-SIGN CRD, obtained at four different saturation times applied into the aliphatic region of the receptor (0 ppm). Taken and adapted from Ref. [21].*

Moving to LDN-DF, the spectrum interpretation was rather simple, as not many STD effects were detected even at longer saturation times (3 sec.). In particular, 6 peaks were only measurable (Figure 5.9A): Fuc_{ter} H2, Fuc_{ter} H4, both Fuc methyl groups and both N-acetyl groups. From the former ones, which were also the most perturbed nuclei, it could be reasonably thought that the terminal Fuc was targeting the primary binding site, since it was the only residue considerably affected by the saturation. In this regard, Fuc_{ter} H1 was likely to be saturated as well, but it cannot be checked out due to the intense overlap with the water signal. Strikingly, the distribution of the remaining STDs resulted particularly strange, as no clear epitopes could be inferred from them. At first, and considering the spectral region in which these groups were located, the possibility that a poorly selective saturation pulse was biasing the experimental observations was taken into account. However, a control STD of the free ligand with the same pulses totally discarded that option, since such a spectrum was completely empty. Therefore, the origin of these effects remained in principle unknown. Overall, the binding epitope of the LDNF antigen involves the entire trisaccharide, whereas that of the LDN-DF tetrasaccharide suggested that only the terminal Fuc establishes direct contacts at the protein surface.

In parallel, the conformation of LDNF was evaluated in the bound state by means of tr-NOESY experiments. Expectedly, as the molecular weight of this compound was ca. 655 Da, the observation of either clear positive or negative NOEs was unpractical at 600 MHz in the free state, as occurred to the blood groups. In this particular case, LDNF displayed slightly negative NOEs at 600

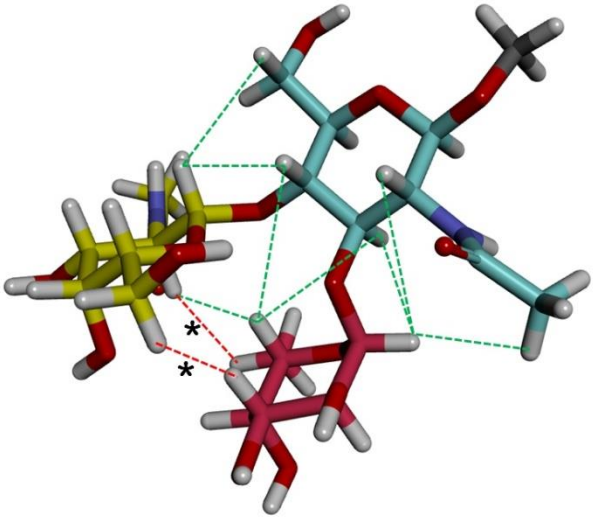
MHz. The spectrum quality marginally improved by increasing the applied field at 800 MHz, but it was not enough for a good characterization. Instead, the trisaccharide was directly modeled through the GLYCAM-Web server [23-24] and the close interresidual distances found on it were simultaneously compared to those noticed in the recorded tr-NOESY and with available data for the Le^x antigen. To clarify, experimental issues relative to low peak intensities hindered the utilization of the 2D-ROESY for the estimation of the free distances.

	GLYCAM model	Le^x Ref. [21]	Tr-NOESY (1:5 CRD/LDNF)
$H1_{Fuc}$-$H2_{Glc}$	3.8	-	2.7-2.8
$H1_{Fuc}$-$H3_{Glc}$	2.6	2.8	2.8-3.0
$H1_{Fuc}$-Ac_{Glc}	3.0	3.8	3.5-3.7
$H5_{Fuc}$-$H3_{Glc}$	3.3	-	-
$H5_{Fuc}$-$H4_{Glc}$	4.0	-	2.5-2.7
$H5_{Fuc}$-$H2_{Gal}$	2.1	-	2.7-2.8
$H1_{Gal}$-$H4_{Glc}$	2.4	2.3	-
$H1_{Gal}$-$H6_{Glc}$	2.7	-	-

Figure 5.10. *Interprotonic distances (Å) directly measured on the GLYCAM-derived model (left column), reported in ref. [21] (medium column), and experimentally estimated through Eq. 2 (Section 4.4), using a tr-NOESY acquired at 298 K with a t_{mix} = 200 ms and 0.2 eq of CRD (right column). The intrarresidual NOEs corresponding to Fuc H1-H2 and Fuc H5-H3 were employed to set two reference distances.*

A global analysis of the collected NOEs (Figure 5.10) led to the preliminary conclusion that the free conformation of the LDNF was being slightly modified upon lectin recognition. In general terms, the GLYCAM model predicted a very packed conformation in solution, with the GalNAc moiety tightly stacked on top of the Fuc ring. This model actually adjusted well to the aforementioned existence of a "non-conventional" hydrogen bond between GalNAc O5 and Fuc H5. However, this conformation placed the Fuc methyl group close to GalNAc H2, and similarly, GalNAc H6 would lie on top of Fuc H4 (Figure 5.11). On the recorded spectrum, none of these two close contacts was detected. Analogously, the Fuc H5/GalNAc H2 NOE was observable, but the interprotonic distance was significantly larger than that in the GLYCAM model. These findings firmly pointed out that the GalNAc and Fuc moieties were not so "stacked" as suggested, but they were more likely to subtly separate from each other during ligand accommodation at the binding site. In either case, it is important to highlight again that these Lewis-type ligands are commonly fairly rigid, showing a limited conformational space [18]. As proof, the other 8

theoretical interresidue contacts that the modeled free structure suggested were actually found in the bound state (Figure 5.12). Unfortunately, signal overlap hampered the estimation of some peak areas, and they are not included in Figure 5.10.



	LDNF	
	φ	Ψ
β-GalNAc	+58.5	+0.0
α-Fuc	+56.5	+20.1

Figure 5.11. *Free conformation of the LDNF antigen, as predicted by GLYCAM06. The dashed green lines represent the close contacts which have been experimentally confirmed to exist in the bound conformation as well. Conversely, the red lines with asterisks denote two interresidual contacts which were not found in the tr-NOESY. The table below includes the corresponding dihedral angles φ and ψ observed for the two non-reducing moieties.*

In view of these results, the GLYCAM-derived model of the free LDNF was straightforwardly used as starting conformation for MD simulations. Both non-reducing sugars displayed glycosidic linkages with *exo-syn* conformations (Figure 5.11), in accordance with the published data for Le^X-containing structures either in solution or by X-Ray [18,25]. Noteworthy, the reported NOEs $H1_{Fuc}$-$H2_{Glc}$ and $H5_{Fuc}$-$H4_{Glc}$ were substantially strong and hence yielded internuclear distances remarkably short, as compared with those present in the modeled antigen. Nevertheless, these values could not be explained on the basis of a torsion on the Fuc-GlcNAc linkage, since the proper geometry of the oligosaccharide implied an opposite effect in these distances: a torsion to

increase one distance inevitably decreased the other one even more. Then, these values were better attributed to simple experimental overestimations in the peak areas.

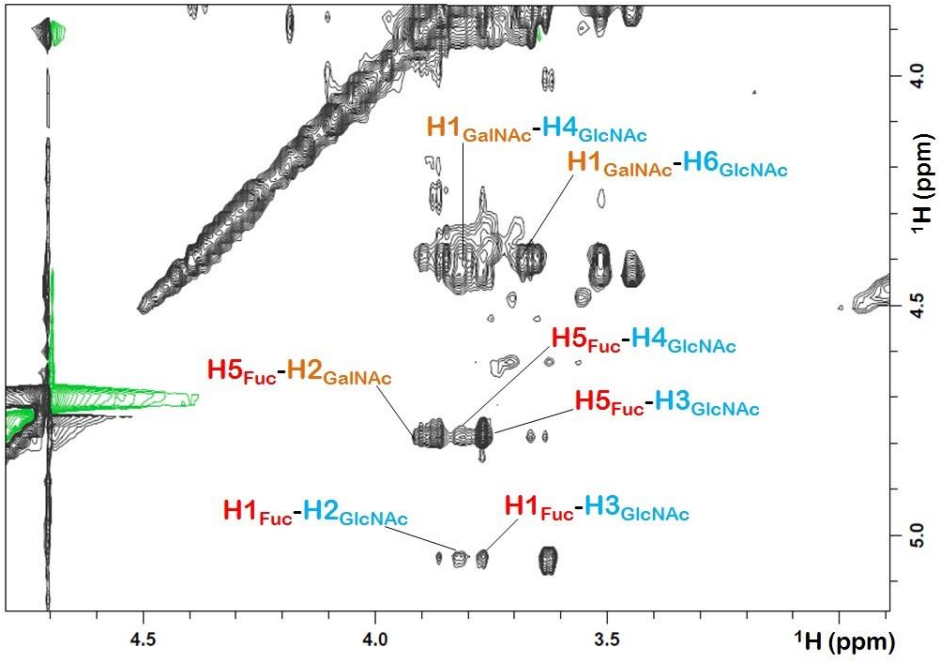


Figure 5.12. *Tr-NOESY spectrum for the LDNF antigen (1 mM ligand, 0.2 mM DC-SIGN CRD, in Tris buffer (D_2O), t_{mix}=200 ms), displaying the eight interresidual contacts detected and assigned.*

Previous to the MD run, the tr-NOESY spectrum was further scrutinized in search of direct interresidue contacts with the lectin, specifically with the Val351 sidechain, which was readily identifiable as explained in *Chapter 4*. Successfully, the methyl groups of Val351 were noticed by means of their direct correlation with Fuc H1 at 0.74 ppm, located in a clean spectral region. Fittingly, following such a correlation led to unequivocally find the intrarresidue NOE with Hα (3.43 ppm), whereas the NOE with Hβ (ca. 1.60-1.80 ppm) was rather complicated to detect due to the extreme peak overlapping. Additionally, two more NOEs appeared at 0.74 ppm in the tr-NOESY spectrum, which corresponded to Fuc H2 and the N-acetyl group from the core GlcNAc (Figure 5.13). Accordingly, the recognition of the Fuc moiety through its OH3 and OH4 groups perfectly fulfilled these observations, and explained reasonably well the STD values experimentally found (Figure 5.14) In particular, the strong STDs of Fuc H1, Fuc H2 and GlcNAc Ac were supported by their direct contacts with the Val351 sidechain. In addition, the weak and uniform STD profile found for the non-reducing GalNAc was in accordance

with its positioning at the secondary site of the CRD defined by strands β_3 and β_4, equivalent to the arrangement seen for the analogous Gal moiety in the Le^X trisaccharide. Under these assumptions, the free ligand geometry (Figure 5.14) was used to carry out the corresponding MD simulations, as help to further evaluate the binding mode.

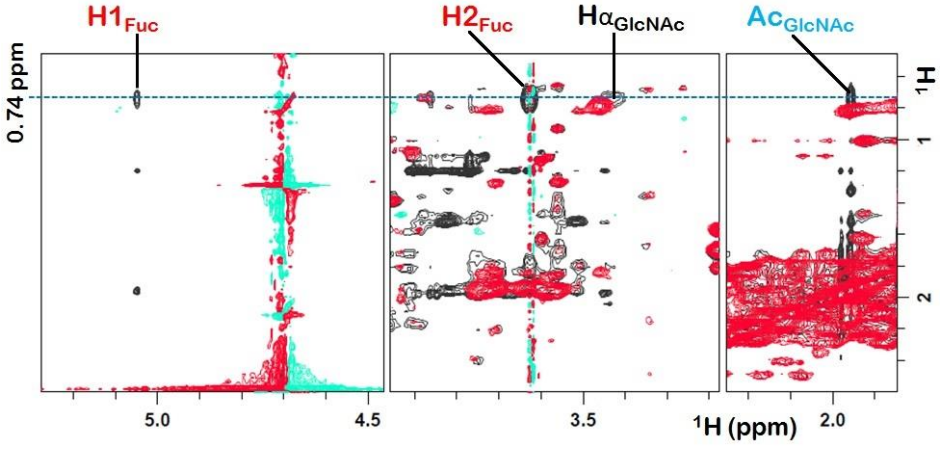


Figure 5.13. *In black, the tr-NOESY of the LDNF-DC-SIGN complex, and in red, the same experiment recorded for the free receptor (0.2 mM). As indicated, some aliphatic correlations appear only in the tr-NOESY which correspond to direct dipolar couplings with the ligand protons.*

Regarding LDN-DF, its low affinity unfortunately precluded the acquisition of the tr-NOESY spectrum, in which the ligand signals were not altered by the presence of DC-SIGN. This is actually indicative, as stated in the *Methods* section, that the k_{off} is rather fast. In this regard, it is worth mentioning that signal recovering observed in *Section 5.1.* was less pronounced in the case of the LDN-DF titration and then it could be feasible that the broadening effects were arising from other unidentified phenomena, for instance non-specific binding, which could occur due to the exaggerate amount of equivalents added. Hence, the STD effects constituted the only source of information, and following this line, the interaction at the calcium site was in principle considered to take place through the external Fuc residue. Also, as already demonstrated by ^{19}F-NMR, the preferential hydroxyl pair to establish this primary interaction is constituted by OH3 and OH4 positions.

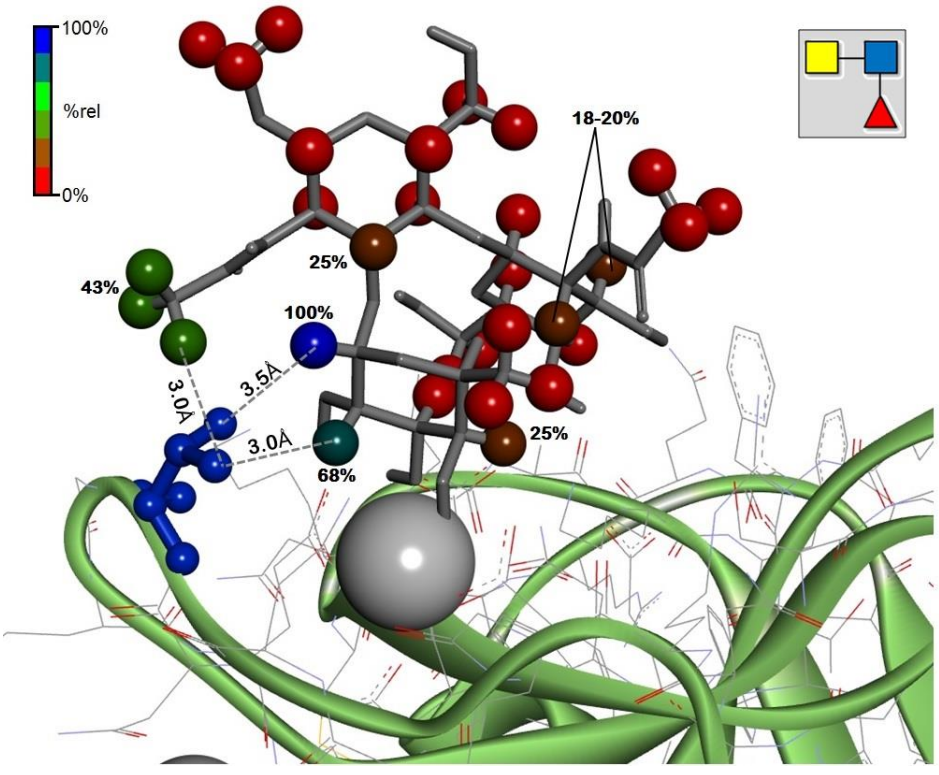


Figure 5.14. *Preliminar binding pose proposed for the LDNF trisaccharide on the basis of the collected STD and tr-NOESY data. Val351 is represented in blue. The protons are colored according to their respective relative STD percentages, from blue (100%) to red (0%). Some important STDs (>20%) are directly labeled. The three transferred NOEs between Val351 methyl groups and the ligand are detailed as well (the protons of the Val351 residue are not represented to simplify the figure).*

5.3. Molecular Dynamics and the 3D structure of the complexes

To eventually achieve a precise description of the recognition events involving the studied ligands and the lectin CRD, the proposed binding poses were subjected to MD simulations. For both oligosaccharides, the starting geometry was built again by using the PDB 1SL5 [25]. For LDNF, the superimposition of its Fuc ring with that of the crystallized ligand led to the pose already described in Figure 5.14, which nicely agreed with the experimental observations. Conversely, the scarce and diffuse information derived from the LDN-DF spectra led to primarily consider more than a unique option. Firstly, the tetrasaccharide was modeled by employing the GLYCAM06 force field as well. In parallel, a test experiment consisting of a 2D-NOESY at low temperature (281 K) was recorded to try verifying the hypothetical

structure of the free ligand in solution. Worth mentioning, the structure of the LDNF trisaccharide "contained" into the LDN-DF scaffold displayed an equivalent arrangement in the modeled ligand. The additional Fuc pyranose attached to the other Fuc seemed to occupy a different region far away from the GlcNAc and the GalNAc residues. Fortunately, such a structure could be in part assessed by the 2D-NOESY, since the only NOEs found for the outer Fuc were those corresponding to intrarresidue contacts (Figure 5.15), being the unique exceptions the expected NOEs between the protons around the glycosidic linkage (Fuc_{ext} H1 and Fuc_{int} H1, H2). Instead, the inner Fuc basically yielded analogous interresidue NOEs as those noticed for the LDNF scaffold. In this scenario, the GLYCAM-derived structure of the LDN-DF tetrasaccharide was analogously employed to build the starting binding pose. Fortunately, neither the superimposition through the inner Fuc nor the docking using the OH2-OH3 pair from the outer Fuc allowed to placing the ligand at the binding site without generating severe clashes, thereby compiling with the starting hypothesis in which the outer Fuc ring was supposed to interact at the binding site through positions OH3 and OH4 (as suggested by the relatively strong STD of Fuc_{ext} H2). Hence, the ligand was again superimposed through the outer Fuc moiety using the LNFP III substrate as template. In both cases, the MD simulations were prolonged for 100 ns and a total of 1000 trajectory points were extracted for the subsequent analyses.

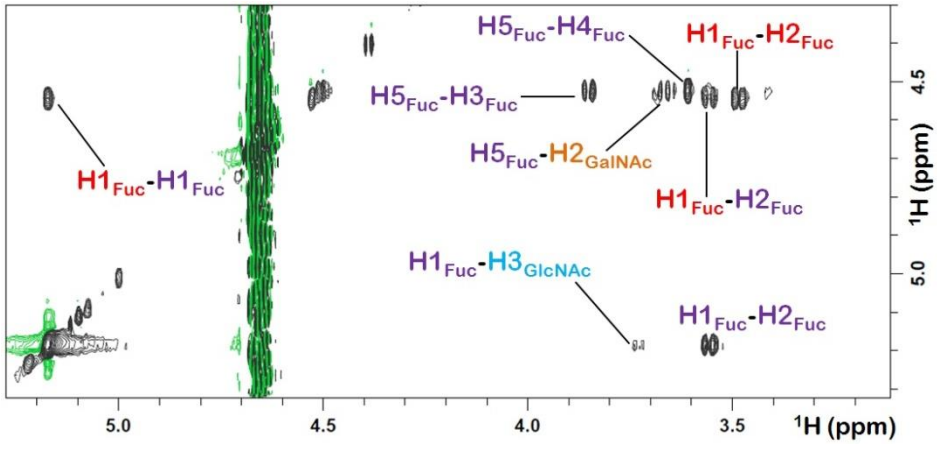


Figure 5.15. *2D-NOESY spectrum of the free LDN-DF antigen partially assigned (T = 281 K, buffer H_2O, t_{mix}=200 ms). Violet labels correspond to the inner Fuc, and red labels to the outer one. As can be noted, the outer Fuc moiety only provided two interresidual NOEs with the neighboring Fuc, in specific with the closest protons to the glycosidic linkage (H1 and H2).*

Successfully, the two oligosaccharides remained anchored to the calcium ion throughout the whole simulation, thus proving the stability of both interactions, regardless the particular affinities. However, the dynamic behaviors were expectedly different between them. From a general perspective, as observed for the blood groups, the LDNF recognition took place through a fairly fixed ligand conformation, which was barely affected or modified along the MD trajectory. Indeed, the dispersion of the ϕ and ψ torsion angles of the two glycosidic bonds involving the core GlcNAc corroborated the rigidity of these ligands (Figure 5.16, *top*). Thus, the ϕ angle for the GalNAc moiety was checked to slightly decrease by 10 degrees with respect to that described in solution, keeping a very low dispersed value around 45-50º. Conversely, the dispersion of ϕ for the Fuc ring was slightly larger (stdev. of ca. 15º), but in any case it ranged around an average value very similar to the angle found in the free model (56º). Equivalent observations were noticed for the corresponding ψ torsions.

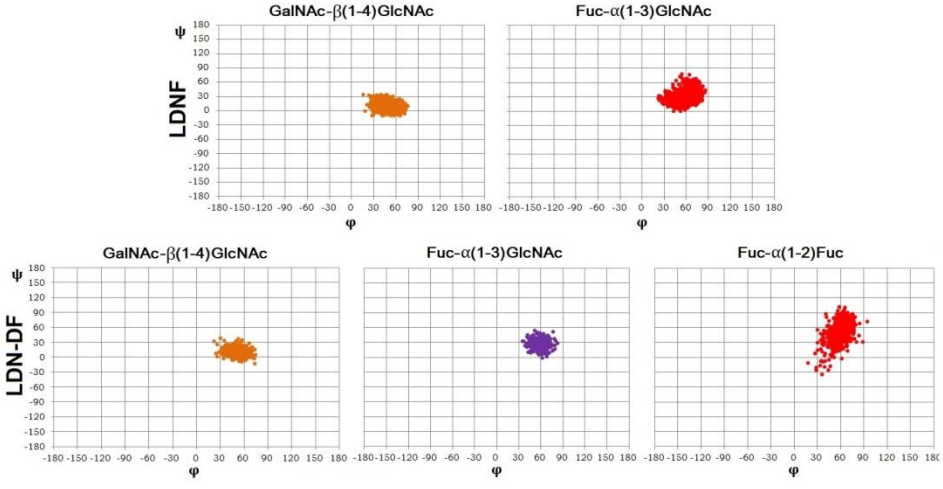


Figure 5.16. *Distribution of the φ and ψ torsional angles which define each glycosidic bond in the studied antigens. The φ dihedral is defined from the anomeric H1 to the opposite carbon (H1-C1-O'-C') and the ψ dihedral, from the anomeric C1 to the opposite proton (C1-O'-C'-H').*

From a global perspective, these subtle variations gave rise to two important observations: firstly, the slight twist of the GalNAc glycosidic angle was enough to noticeably separate this residue from the stacking Fuc, provoking an increase in the Fuc H5–GalNAc H2 distance of ca. 1 Å in average (2.8-3.0 Å), which fitted very well with the larger internuclear distance derived from tr-NOESY. In addition, this small change allowed the GalNAc ring to perfectly accommodate on the extended binding site of the DC-SIGN CRD, where other

ligands have been already reported to interact [26-27]. Given that the Fuc ring was tightly attached to the metal ion through both the coordination bonds and the additional non-covalent bonds (H-bonds) with the surrounding sidechains (Figure 5.17, *left*), the mobility of the two glycosidic linkages exclusively affected the non-reducing GalNAc, which showed a more diffused position during the simulation. Specifically, this pyranose was observed to move in parallel respect to the β-strands below, establishing in all cases stable non-covalent interactions with three nearby protein residues: E358, S360 and D367 (Figure 5.18). These interactions were actually formed in two ways: simultaneously targeting both S360 and E358 polar groups through OH6, or alternatively providing three different H-bonds which in some cases were formed at the same time: OH6-D367 (the major one), OH4-E358, and OH3-S360. The interplay between these different modes of interaction constantly occurred along the MD run, as verified by monitoring the distance from some surrounding amino acids to the cited ring (Figure 5.19). As remark, it was clear that there were always stable secondary interactions provided by this residue, with position 6 being especially crucial.

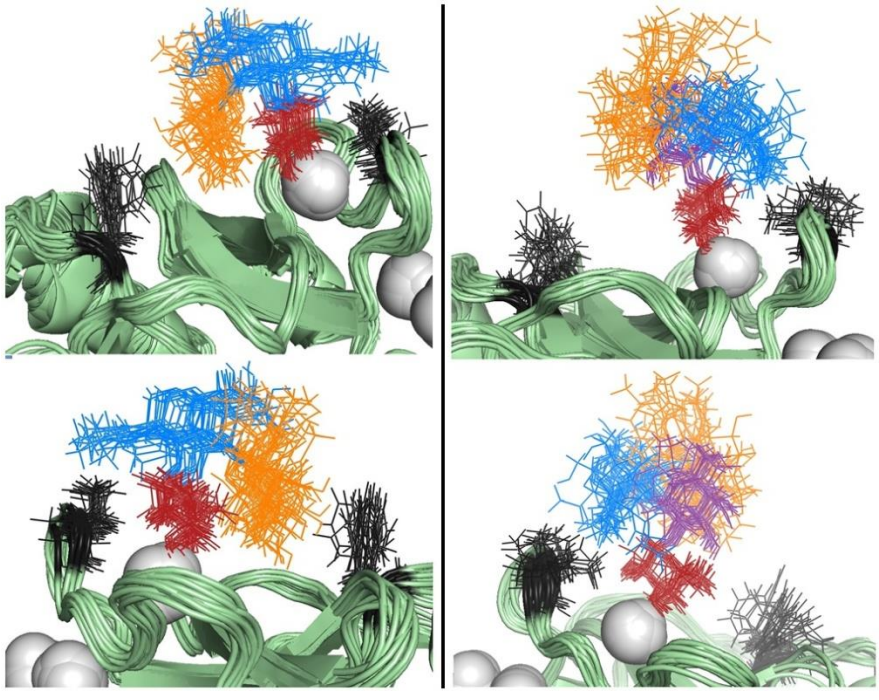


Figure 5.17. *Ligand trajectories observed along the MD simulations (left: LDNF; right: LDN-DF). The bundle is constituted by 20 representative frames. Color code: blue: β-GlcNAc; purple: inner Fuc; red: outer Fuc; orange: β-GalNAc. The white spheres represent the calcium ions. Residues Val351 and Phe313 are detailed as black lines.*

Finally, it is also worth mentioning the close interaction with the protein provided by the N-acetyl group of the central GlcNAc. In agreement with the experimental data from STD and tr-NOESY, this acetyl group stacks on top of the Val351 sidechain, directly contacting at least one of the two methyl groups of this amino acid (~3 Å, Figure 5.19). This result specially underlined the crucial role of such a substituent in the stabilization of the Lewis-type antigens when interacting with DC-SIGN, and strongly supports the negligible binding observed for other related lectins, as L-SIGN [28].

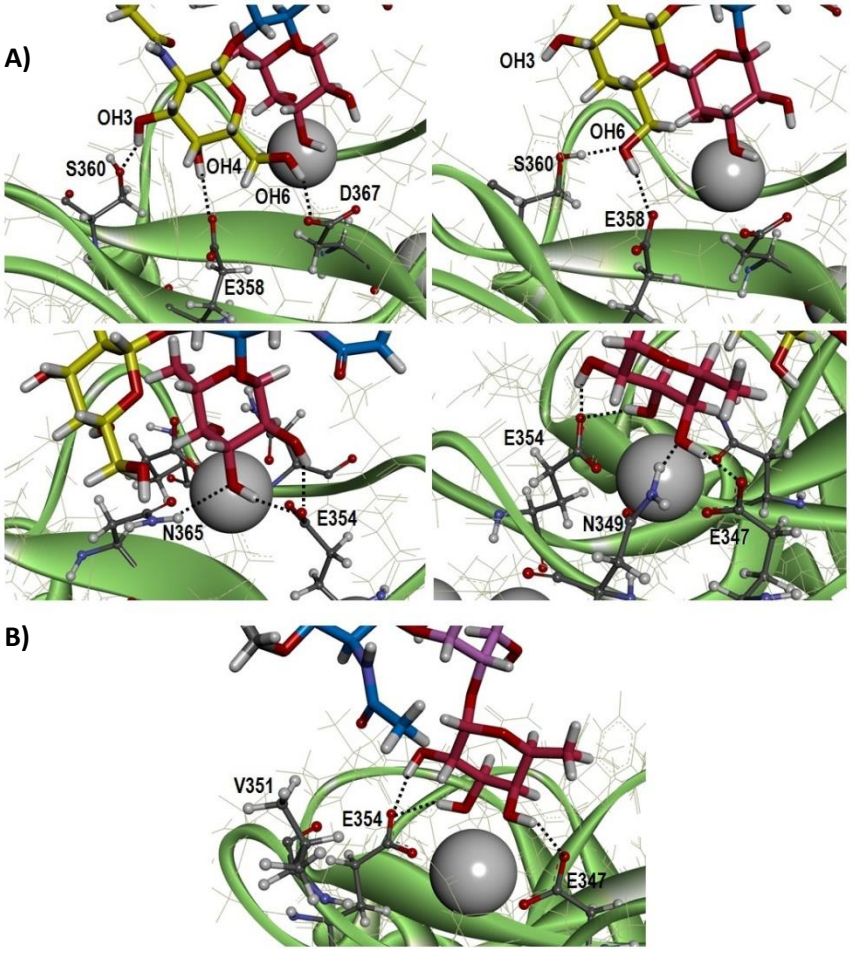


Figure 5.18. *A) On the top row, detailed analysis of the molecular interactions (H bonds, dashed black lines) involving the non-reducing GalNAc moiety of the LDNF trisaccharide. On the bottom row, front and back views of the binding site displaying the molecular interactions provided by the Fuc residue. The coordination bonds are not represented. B) H-bond network established by the non-reducing Fuc in the LDN-DF complex. As noted, the residue Val351 lies fairly away from the Fuc ring.*

The scenario was totally different for LDN-DF, which certainly agreed with the poor ability exhibited by this ligand for DC-SIGN targeting. The glycosidic linkages involving the core GlcNAc marginally varied, with even less dispersion than those of LDNF. In deep contrast, the Fucα1-2Fuc linkage presented a relative higher dispersion of the angle values, especially concerning ψ (Figure 5.16, *bottom*). Again, taking into account the network of coordination bonds and H-bonds which fixed the relative position of the interacting outer Fuc, the high angle variability acutely affected to the entire LDNF trisaccharide placed above, which displayed a considerably undefined position along the MD trajectory (Figure 5.17, *right*). In addition, the particular geometry of the complex inevitably led the upper trisaccharide scaffold to always stay far away from the protein surface. Thus, no stable contacts with the protein arising from these residues were noticed at all, and consequently, the interacting epitope was exclusively reduced to the single non-reducing Fuc (Figure 5.18). In essence, the few experimental data supported these results as well, since no other binding poses would explain the STDs from Fuc_{ext} H2 and Fuc_{ext} H4. Moreover, the affinity contributed by a single Fuc is expectedly low as well as the entropic penalty will be larger for a tetrasaccharide than for a monosaccharide.

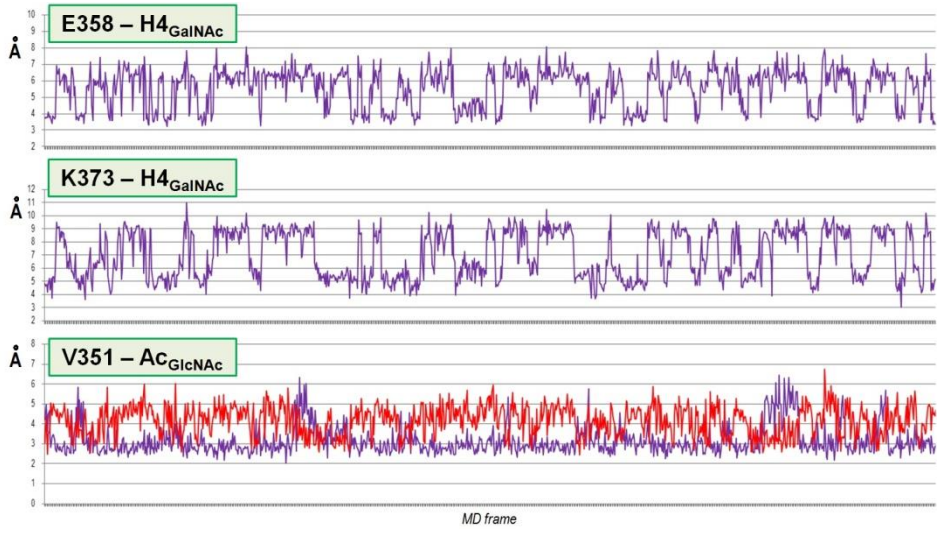


Figure 5.19. *Monitoring of the intermolecular distances between the proximal sugar residues of LDNF and some residues of the protein along the MD simulation. Top and middle: distance from $H4_{GalNAc}$ to the terminal oxygen and nitrogen of E358 and K373, respectively. As noted, both distances oscillate around two average values. Bottom: distance from the Ac_{GlcNAc} protons to both methyl groups from the V351 sidechain.*

To eventually assess the veracity of the proposed 3D models, the average theoretical STD values were estimated for both ligand trajectories by CORCEMA-ST calculations. As the particular conditions applied for the STD experiments were essentially equivalent as those used for the blood groups (saturation, concentrations of protein and ligands), the same theoretical parameters were taken to prepare the CORCEMA setup. As exceptions, the particular K_d (1 mM) and k_{on} (10^5 M^{-1} s^{-1}) constants estimated for the LDNF antigen from the HSQC-based titrations were applied for the analysis of this trisaccharide. Unfortunately, these two values remained unknown for the other antigen. As potential solution, the same k_{on} rate was supposed under the assumption that the LDN-DF affinity was about 10 mM, considering the results from NMR titrations. These values would mean that the off-rate ranges around 10^3 s^{-1}, which is equivalent to consider a fast-intermediate exchange regime for this ligand, partially in accordance with the deductions presented so far.

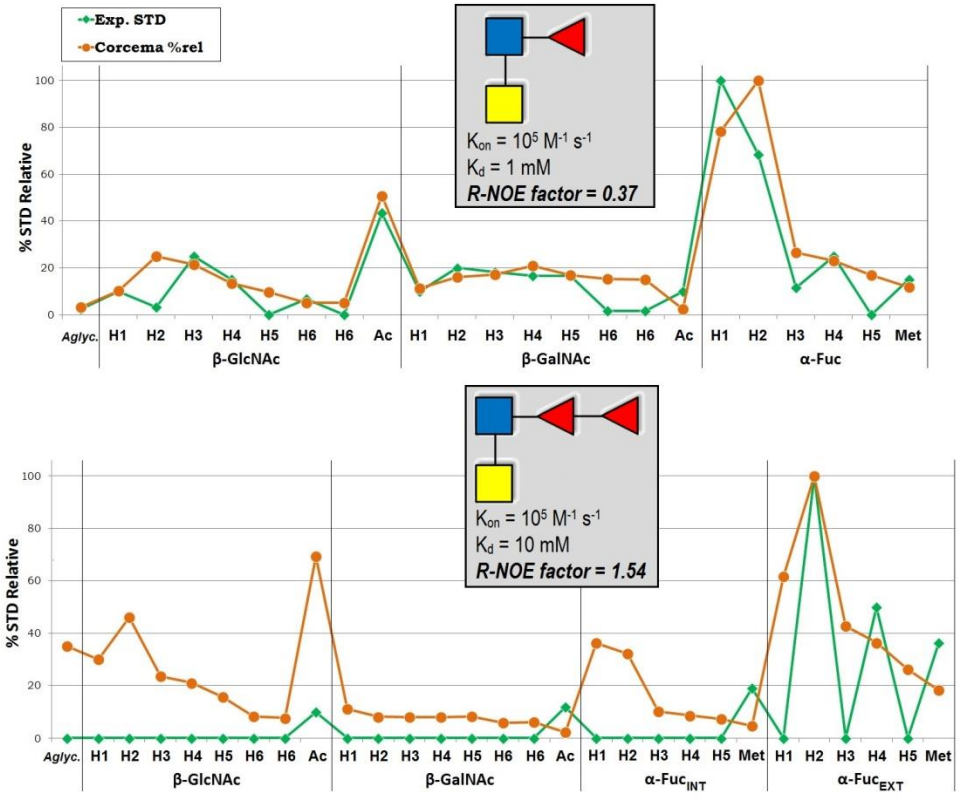


Figure 5.20. *CORCEMA-ST-derived STDs (orange) as compared with the experimental ones (green). The theoretical values are also expressed in relative scale (Ref. H2 Fuc in both cases). CORCEMA setup: Δ_{sat}: Ile, Leu, Val, Ala, Thr; t_{sat} = 2 sec.; $[P_0]$ = 60 µM; $[L_0]$ = 4.2 mM; rholeak = 0.1; B_0 = 600 MHz; τ_{free} = 0.5 ns; τ_{bound} = 8.33 ns; τ_{intra} = 0.01 ns.*

The calculations were performed for 100 MD frames randomly selected in each case. As already stated, this amount is enough to achieve representative results as long as no unrealistic conformations or binding poses are considered. Successfully, the CORCEMA profile obtained for LDNF matched substantially well with the experimental STD pattern, only displaying minor discrepancies (Figure 5.20). Indeed, the only detail that deserves attention is the fact that CORCEMA-ST predicted the Fuc H2 nucleus to be the most affected rather than Fuc H1, as experimentally found. As already stated, such a prediction better fitted with the relative values reported for the blood groups in this work and also for the Le^X antigen. Moreover, the Fuc ring always displays Fuc H2 closer to Val351 in the MD simulations, both for the blood group antigens and for the LDNF. At this point, considering this, the exact reason why the STD of Fuc H1 was more intense than that of Fuc H2 remained a mystery. In either case, the R-NOE factor was likewise good considering that no further optimizations were applied either to the CORCEMA parameters or to the MD simulations (e.g. restrictions via experimental parameters) [29-30]. Therefore, from an overall perspective, the proposed 3D model for the DC-SIGN-LDNF complex can be considered as robust, since it is well justified by the NMR observations and further supported by computational calculations (Figure 5.21).

Regarding the LDN-DF antigen, the CORCEMA profile expectedly yielded a poor correlation between the theoretical values and the limited experimental data (Figure 5.20). In fact, the latter aspect likewise limited the discussion of the results, as no realistic comparisons could be done given the lack of information derived from the extremely low ligand affinity. As highlights, the CORCEMA-ST interestingly predicted the saturation profile of the recognized Fuc to be similar to that observed for the LDNF. Also, simulations suggested a certain weak and uniform saturation of the GlcNAc residue, and intriguingly, a stronger saturation of the N-acetyl group than that observed in the acquired STD. Certainly, the GlcNAc moiety was observed to move around the region above Val351, but this movement was rather erratic and distant, and in no case generated close and stable contacts with the cited residue. Even receiving saturation as a result of this “overflying”, the direct involvement of this moiety in the recognition event is fairly unlikely. In conclusion, the proposed 3D model for the LDN-DF ligand served to explain its markedly low affinity, as checked out by NMR methods. However, although it can be considered as valid for now, the scarce information from the ligand perspective inevitably left the question about the molecular recognition of this antigen partially opened.

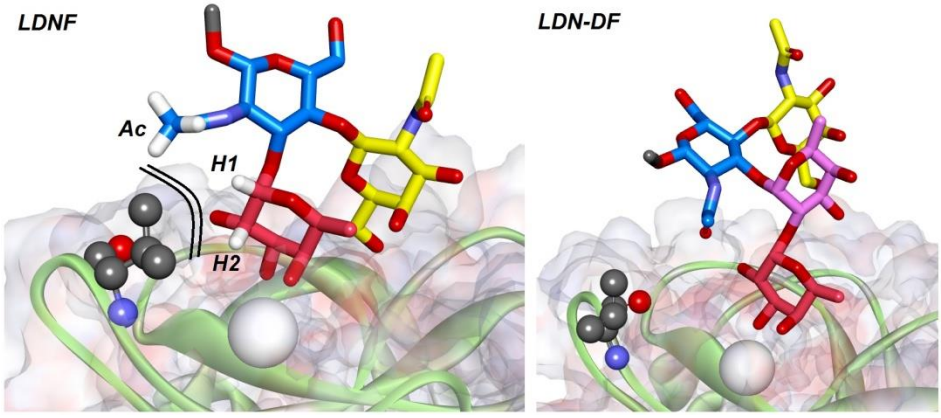


Figure 5.21. *Final CRD-antigen complexes proposed for the studied ligands (Left: LDNF; right: LDN-DF). Color code: Red: non-reducing α-Fuc. Pink: inner α-Fuc. Blue: β-GlcNAc. Yellow: β-GalNAc. The white spheres represent the calcium atoms. The Val351 sidechain is detailed, as well as the hydrophobic stabilization provided by Fuc H1, Fuc H2 and Ac GlcNAc in the LDNF-CRD complex.*

In conclusion, the structural model for LDNF was very similar to that described for the Le^X trisaccharide, indicating that the presence of a β4-linked GalNAc vs. a β4-linked Gal moiety does not substantially alter such binding.

The terminal α1,2-linked fucoside of the LDN-DF tetrasaccharide can also bind into the canonical binding site of DC-SIGN but in this case, the GlcNAc and GalNAc residues are placed away from the protein surface preventing additional contacts. As a result, the binding affinity is at least an order of magnitude lower. Interestingly, structures such as LDN-DF and LDNF are part of complex oligosaccharides in which multiple copies of these epitopes can be presented. Potentially, such structures can display multivalent interactions leading to high avidity of binding. Further biological studies with these antigens using multiple presentations are currently underway in the laboratory of Prof. G.-J. Boons at Utrecht.

5.4. References

1. P. T. LoVerde. Schistosomiasis. *Adv. Exp. Med. Biol.,* (2019), **1154**, 45-70.
2. P. Kalantari, S. C. Bunnell and M. J. Stadecker. The C-type lectin receptor-driven, Th17 cell-mediated severe pathology in schistosomiasis: not all immune responses to helminth parasites are Th2 dominated. *Front. Immunol.,* (2019), **10**, 26.

3. M. L. Mickum, N. S. Prasanphanich, J. Heimburg-Molinaro, K. E. Leon and R. D. Cummings. Deciphering the glycogenome of schistosomes. *Front. Genet.,* (2014), **5**, 262.
4. M. J. H. Meevissen, N. N. Driessen, H. H. Smits, R. Versteegh, S. J. van Vliet, Y. van Kooyk, G. Schramm, A. M. Deelder, H. Haas, M. Yazdanbakhsh and C. H. Hokke. Specific glycan elements determine differential binding of individual egg glycoproteins of the human parasite *Schistosoma mansoni* by host C-type lectin receptors. *Int. J. Parasitol.,* (2012), **42**, 269-277.
5. C. H. Hokke, A. M. Deelder, K. F. Hoffmann and M. Wuhrer. Glycomics-driven discoveries in schistosome research. *Exp. Parasitol.,* (2007), **117**, 275-283.
6. A. van Remoortere, C. H. Hokke, G. J. van Dam, I. van Die, A. M. Deelder and D. H. van den Eijnden. Various stages of schistosoma express Lewis(x), LacdiNAc, GalNAcbeta1-4 (Fucalpha1-3)GlcNAc and GalNAcbeta1-4(Fucalpha1-2Fucalpha1-3)GlcNAc carbohydrate epitopes: detection with monoclonal antibodies that are characterized by enzymatically synthesized neoglycoproteins. *Glycobiology,* (2000), **10**, 601-609.
7. I. van Die and R. D. Cummings. Glycan gimmickry by parasitic helminths: a strategy for modulating the host immune response? *Glycobiology,* (2010), **20**, 2–12.
8. N. A. Peterson, T. K. Anderson and T. P. Yoshino. *In silico* analysis of the fucosylation-associated genome of the human blood fluke *Schistosoma mansoni*: cloning and characterization of the fucosyltransferase multigene family. *PLoS ONE,* (2013), **8**, e63299.
9. S. Frank, I. vanDie and R. Geyer. Structural characterization of *Schistosoma mansoni* adult worm glycosphingolipids reveals pronounced differences with those of cercariae. *Glycobiology,* (2012), **22**, 676–695.
10. A. Dell, H. R. Morris, R. L. Easton, M. Panico, M. Patankar, S. Oehninger, R. Koistinen, H. Koistinen, M. Seppala and G. F. Clark. Structural analysis of the oligosaccharides derived from glycodelin, a human glycoprotein with potent immunosuppressive and contraceptive activities. *J. Biol.Chem.,* (1995), **270**, 24116–24126.
11. C. W. A. Naus, A. van Remoortere, J. H. Ouma, G. Kimani, D. W. Dunne, J. P. Kamerling, A. M. Deelder and C. H. Hokke. Specific antibody responses to three schistosome-related carbohydrate structures in recently exposed immigrants and established residents in an area of *Schistosoma mansoni* endemicity. *Infect. Immun.,* (2003), **71**, 5676–5681.
12. I. van Die, S. J. van Vliet, A. K. Nyame, R. D. Cummings, C. M. C. Bank, B. Appelmelk, T. B. Geijtenbeek and Y. van Kooyk. The dendritic cell-specific C-type lectin DC-SIGN is a receptor for *Schistosoma mansoni* egg antigens and recognizes the glycan antigen Lewis x. *Glycobiology,* (2003), **13**, 471-478.
13. H. Wegener, A. Mallagaray, T. Schöne, T. Peters, J. Lockhauserbäumer, H. Yan, C. Uetrecht, G.S. Hansman and S. Taube. Human norovirus GII.4(MI001) P dimer

binds fucosylated and sialylated carbohydrates. *Glycobiology,* (2017), **27**, 1027–1037.

14. A. Mallagaray, C. Rademacher, F. Parra, G. Hansman and T. Peters. Saturation transfer difference nuclear magnetic resonance titrations reveal complex multistepbinding of L-fucose to norovirus particles *Glycobiology,* (2017), **27**, 80–86.
15. M. P. Williamson. Using chemical shift perturbation to characterize ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.,* (2013), **73**, 1-16.
16. G. Timpano, G. Tabarani, M. Anderluh, D. Invernizzi, F. Vasile, D. Potenza, P. M. Nieto, J. Rojo, F. Fieschi and A. Bernardi. Synthesis of novel DC-SIGN ligands with an α-fucosylamide anchor. *ChemBioChem,* (2008), **9**, 1921–1930.
17. J. Feeney, J. G. Batchelor, J. P. Albrand and G. C. K. Roberts. Effects of intermediate exchange processes on the estimation of equilibrium constants by NMR, *J. Magn. Reson.,* (1979), **33**, 519-529.
18. M. D. Battistel, H. F. Azurmendi, M. Frank and D. I. Freedberg. Uncovering nonconventional and conventional hydrogen bonds in Oligosaccharides through NMR experiments and molecular modeling: application to Sialyl Lewis-X. *J. Am. Chem. Soc.,* (2015), **137**, 13444−13447.
19. P. Valverde, J. I. Quintana, J. I. Santos, A. Ardá, and J. Jiménez-Barbero. Novel NMR avenues to explore the conformation and interactions of glycans. *ACS Omega.,* (2019), **4**, 13618-13630.
20. T. Aeschbacher, M. Zierke, M. Smieško, M. Collot, J. M. Mallet, B. Ernst, F. Allain and M. A. Schubert. Secondary structural element in a wide range of fucosylated glycoepitopes. *Chem. Eur. J.,* (2017), **23**, 11598−11610.
21. K. Pederson, D. A. Mitchell, and J. H. Prestegard. Structural characterization of the DC-SIGN−LewisX complex. *Biochemistry,* (2014), **53**, 5700−5709.
22. B. Meyer and T. Peters. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew. Chem. Int. Ed. Engl.,* (2003), **42**, 864-890.
23. K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.,* (2008), **29**, 622-655.
24. B. L. Foley, M. B. Tessier and R. J. Woods. Carbohydrate force fields. *Wiley Interdiscip. Rev. Comput. Mol. Sci.,* (2012), **2**, 652-697.
25. Y. Guo, H. Feinberg, E. Conroy, D. A. Mitchell, R. Alvarez, O. Blixt, M. E. Taylor, W. I. Weis and K. Drickamer. Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat. Struct. Mol. Biol.,* (2004), **11**, 591−598.
26. H. Feinberg, R. Castelli, K. Drickamer, P. H. Seeberger and W. I. Weis. Multiple modes of binding enhance the affinity of DC-SIGN for high mannose N-linked glycans found on viral glycoproteins. *J. Biol. Chem.,* (2007), **282**, 4202−4209.

27. H. Feinberg, D. A. Mitchell, K. Drickamer and W. I. Weis. Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science,* (2001), **294**, 2163–2166.
28. E. van Liempt, A. Imberty, C. M. C. Bank, S. J. van Vliet, Y. van Kooyk, T. B. Geijtenbeek and I. van Die. Molecular basis of the differences in binding properties of the highly related C-type lectins DC-SIGN and L-SIGN to Lewis X trisaccharide and *Schistosoma mansoni* egg antigens*. J. Biol. Chem.,* (2004), **279**, 33161-33167.
29. Y. Yuan, D. W. Bleile, X. Wen, D. A. Sanders, K. Itoh, H. W. Liu and B. M. Pinto. Investigation of binding of UDP-Galf and UDP-[3-F]Galf to UDP-galactopyranose mutase by STD-NMR spectroscopy, molecular dynamics, and CORCEMA-ST calculations. *J. Am. Chem. Soc.,* (2008), **130**, 3157–3168.
30. N. R. Krishna and V. Jayalakshmi. Complete Relaxation and Conformational Exchange Matrix Analysis of STD-NMR spectra of ligand receptor complexes. *Prog. Nucl. Magn. Reson. Spectrosc.,* (2006), **49**, 1-25.

CHAPTER 6: SHORT TERM STAY: DE NOVO SYNTHESIS OF 3-DEOXY-3-FLUOROFUCOSE

The following chapter gathers the results derived from the collaboration carried out during the short term stay (4 months) at the University of Southampton. With the purpose of evaluating the structural requirements that makes the Fuc moiety the strongest DC-SIGN binder, three derivatives were chosen (2F-Fuc, 3F-Fuc and 4F-Fuc) for the T_2 filtered ^{19}F NMR experiments, considering that the natural interaction sugar/lectin should take place through hydroxyls at positions 2, 3 and/or 4. Of these three compounds, the 3F-Fuc was the only structure whose synthesis had been never accomplished before. Therefore, the goal for the stay was to afford it by designing an efficient *de novo* chemical synthesis.

6.1. Introduction

Chemical synthesis of fluorine-containing carbohydrates has been extensively reported over the last decades [1]. Nowadays, it is still an active area of research, with many groups working on developing new and more efficient strategies to overcome the usual difficulties associated with functionalizing sugars, mainly those regarding regio- and stereoselectivity [2]. The presence of multiple chiral centers and several positions with similar reactivity usually implies the incorporation of protective steps and the potential formation of many side products [3]. Most of the available published data on sugar fluorination covers the functionalization of the common monosaccharides: D-glucose [4], D-galactose [5], and D-mannose [6], as well as D-pentoses [7]. Currently, Fuc is known to be a fundamental brick on the structure of N- and O-linked glycans and glycolipids present in mammalian cells [8], where it is commonly attached to the terminal positions, taking part of biologically relevant antigens such as the histo-blood groups and Lewis-type motifs [9], which are ligands for specific protein receptors [10-13]. Hence, the increasing interest on this sugar and its potential use in therapeutics has encouraged the synthesis of new L-fucose mimetics over the last decades, including fluorinated derivatives.

2-deoxy-2-fluorofucose (2F-Fuc) was actually the first fluorinated analogue of this monosaccharide to be prepared, and many optimized synthetic routes have been proposed since then [14-18]. In parallel, **fluorination at position C6** was also described by directly reacting a protected L-Gal derivative with DAST [19], and later, an improved method to yield the corresponding trifluoromethyl derivative starting from D-lyxose was presented [20]. Both 2-fluoro and 6,6,6-trifluoro Fuc analogues immediately gained much more attention for their capability to inhibit the enzymatic fucosylation of

glycoproteins [21-23]. This particular property, in the case of the 2F-Fuc, has been widely investigated for the development of antitumoral therapies [24-26] and treatments for other diseases, like the sickle cell disease (SCD) [27]. 2F-Fuc is commercially available and very recently, a manufacturing process have been designed to commercialize the 6,6,6-trifluorofucose analogue as well [28]. Functionalization at position C4 has been attempted by different methods, most of them based on the previous preparation of partially protected L-Fuc derivatives, readily affordable given the lower reactivity of the axial OH-4 [29]. However, the specific introduction of a fluorine atom by nucleophilic substitution was not straightforward due to the poor nucleophilic character of the fluoride anion. Very recently, an approach based on the previous inversion of the 4-OH configuration of the protected L-Fuc has permitted the synthesis of the partially protected L-quinovose (6-deoxy-L-Glc), which yielded **4-deoxy-4-fluorofucose** (4F-Fuc) upon treatment with DAST [30]. Regarding our target, **3-deoxy-3-fluorofucose** (3F-Fuc), no reports were found. Therefore, a new synthetic route was planned based on analogous processes that successfully enabled the attachment of a fluorine atom at C3 in similar sugar scaffolds.

6.2. Preliminary tests employing classical synthetic approaches

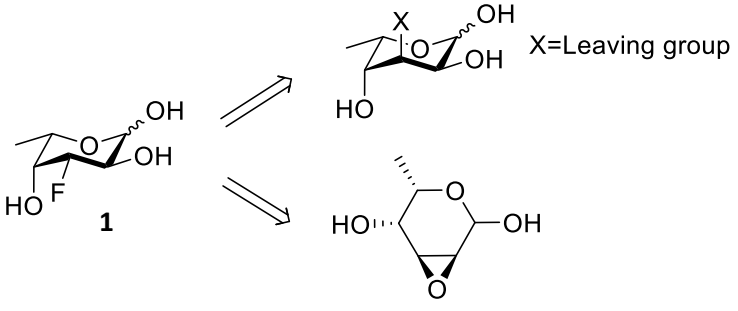


***Figure 6.1**. Basic retrosynthetic scheme including the two synthetic strategies designed to get **1**, starting from a cyclic scaffold with the suitable stereochemistry.*

From a general perspective, the preparation of pyranose analogues unprotected at O3 remains challenging. The nucleophilic aperture of epoxides and aziridines at positions C2 and C3 is often limited by regioselectivity problems, leading to mixtures of products with low overall yields [31-32]. Similarly, the direct S_N2 substitution of leaving groups is hampered by the weak nucleophilicity of fluoride. This is also translated into low yields and complex mixtures resulting from side elimination reactions [33] and

substitutions driven by the proper solvent species or other neighboring nucleophilic groups in the scaffold [34]. Despite these drawbacks, there are successful examples on selective fluorination at position 3 of sugars, although all of them highlight the importance of thoroughly evaluating the stereochemistry of the specific substrate, the presence of additional reactive groups and the reactants and reaction conditions [35-38]. In this scenario, the *de novo* synthesis of 3-deoxy-3-fluoro-L-Fuc was planned by considering the simplest options mentioned above: the aperture of an epoxide moiety and the nucleophilic substitution of a leaving group (Figure 6.1).

6.2.1. Substitution of a leaving group

The direct S_N2 substitution is not straightforward and entails serious complications to functionalize the 3F-Fuc ring at position 3, as has been recently exemplified by Shimabukuro et al [46] (Figure 6.2A). Although the introduction of a leaving group at C3 is readily affordable, the stereochemistry at C4 enables the existence of β-trans-axial effects [47], arising from electron withdrawing groups placed at that position (Figure 6.2B). Even so, functionalization at C3 with a triflate group has been successfully carried out, although with many protection and deprotection steps, as shown in Figure 6.2A. Thus, the S_N2 alternative was first discarded and other less time-consuming routes were sought.

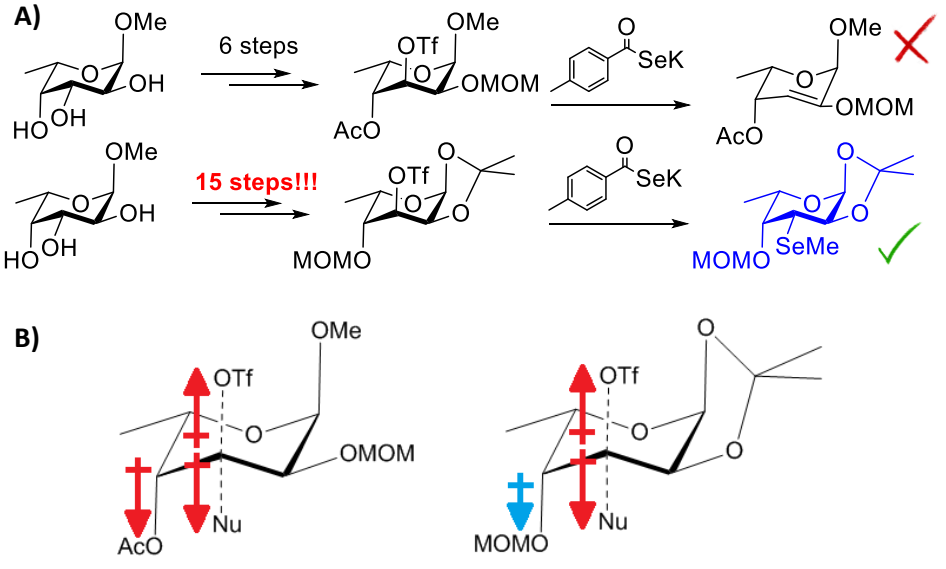


Figure 6.2. *A) The two synthetic routes employed to afford 3-seleno-fucose derivatives [46]. On the top, the failed route using acetate as protecting group at position 4. On the bottom, the alternative strategy to build the same triflate derivative with a*

methoxymethyl ether group (MOM) placed at position 4. B) Schematic representation of the β-trans-axial effect existing in the Fuc ring as a result of the nucleophilic attack at position 3 (S_N2 transition state). In both cases, the neighboring electron withdrawing group gives rise to a repulsive alignment between the fixed dipole at C4 and the new dipole that is being formed at the adjacent carbon. The permanent dipole created by the presence of the acetyl group is stronger than that of the MOM, acutely hindering the nucleophilic displacement.

6.2.2. Epoxide opening

A nucleophilic attack on an epoxide always takes place in a stereospecific fashion, such that the final product necessarily displays the neighboring oxygen in an *anti-type* configuration. Then, for our particular case, the stereochemistry of the final product demands the formation of the epoxide exclusively at positions 2 and 3, with the oxygen atom facing to the top part of the ring. However, according to the already described epoxide opening for α–methyl-L-fucopyranosides [39], the expected regioselectivity would be opposed to the desired one [40] (Figure 6.3).

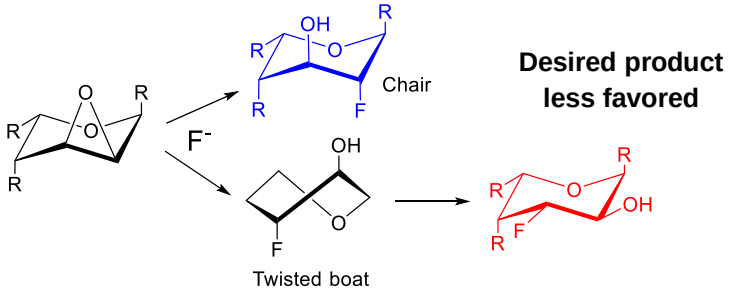


Figure 6.3. *Representation of the C2-C3 epoxide stereochemistry when undergoing the two possible fluoride-mediated opening processes. The case depicted below, which leads to the final compound, shows the unfavorable transition conformation adopted, just before reaching the more stable $^1C_4$ chair.*

A potential solution to overcome this drawback could be achieved using the same strategy in furanoses rather than in pyranoses. In fact, the regioselective opening of 2,3-epoxy-furanoses at C3 has been already reported starting from protected D-xylose [41-42] and D-arabinose [43-44] analogues to form the correct epoxide in few steps (Figure 6.4A). In these cases, the KHF_2/NaF mixture was always used as nucleophilic species. Intriguingly, there is one reported case of an epoxide opening in a β-methyl-L-fucofuranoside scaffold that exclusively lead to the C2-functionalized product [45] (Figure 6.4B). As the nucleophile was completely different in the latter case (hydride), there were no evidences to ensure *a priori* whether the reaction would work with

fluoride. Moreover, the access to the 2,3-epoxy-L-fucofuranoside substrate would involve 7-10 synthetic steps starting from the commercial 1,2-isopropylidene-D-glucofuranose.

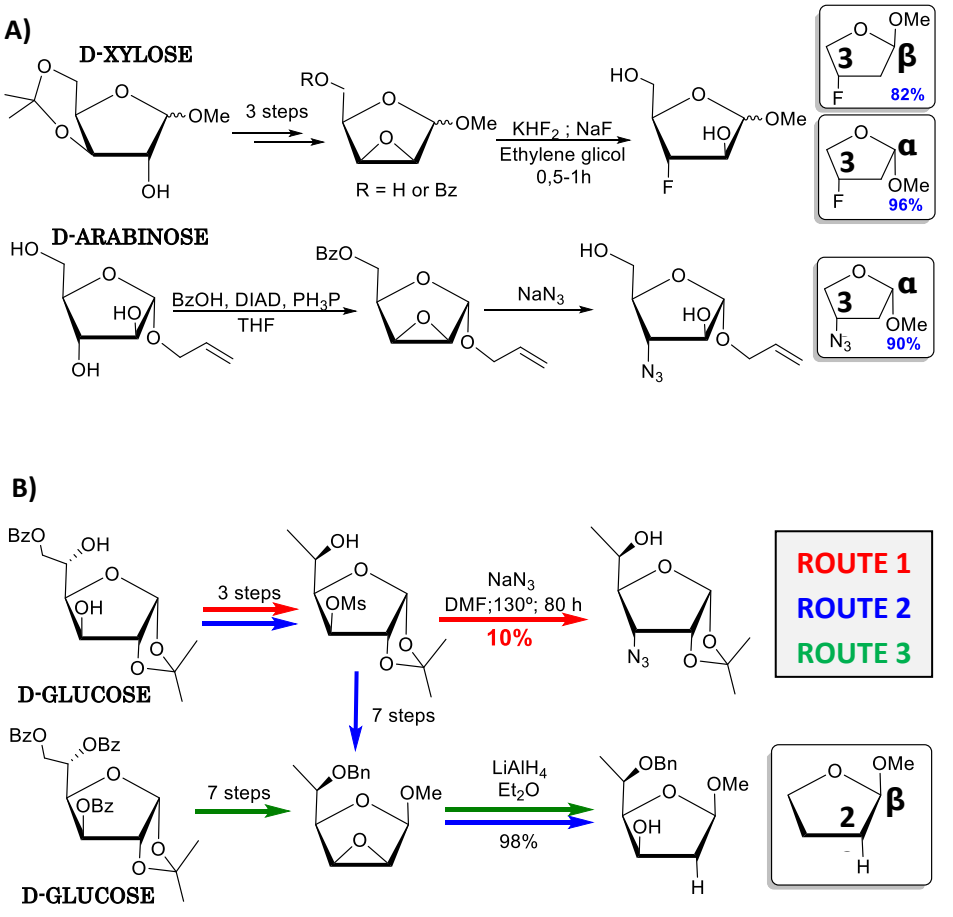


Figure 6.4. *Comparison of the described synthetic methodologies to regioselectively functionalize 2,3-epoxyfuranoses. A) Fluorination at position 3 achieved in two different pentoses [41,43]. B) Different routes to transform the D-glucofuranose into L-furanoses functionalized at either C2 or C3 [45].*

The 3-mesylate derivative is easier to afford (route 1 in red), but the yields would be notably lower due to the equatorial arrangement unfavorable for the S_N2 attack. Moreover, the final product would not display the L-Fuc stereochemistry and further transformations would be required. Fortunately, a shortcut for that problem could be envisaged, based on a stannylene-mediated benzylation procedure recently published [48], which would permit to selectively prepare the protected β-L-fucofuranoside directly from commercial L-Fuc. The corresponding 2,3-epoxy-L-fucofuranoside could be then afforded in only two steps following the methodology applied for D-

arabinose, as shown in Figure 6.5. Counting on the subsequent fluorination and deprotection steps, the global synthesis could be presumably addressed in 4-5 steps.

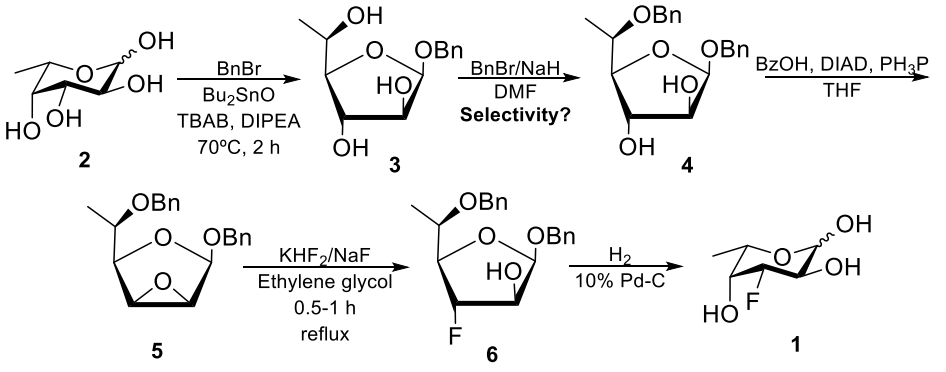


Figure 6.5. *Proposed plan to synthesize the 3-fluorofucose from commercial L-fucose in three main steps (formation of the protected furanose, formation of the epoxide and regioselective epoxide opening), combining the methodologies described in Refs. [42,43,48].*

6.2.3. Results

For the first step, the commercial mixture of α/β-Fuc was reacted with an excess of benzyl bromide in the presence of sub-stoichiometric amounts (0.1 eq) of dibutyltin oxide, following the specific conditions described in Ref. [48] (Figure 6.6). The major product of the reaction was the monobenzylated sugar, as checked out by HRMS. In addition, the chemical shift of H1 (5.1 ppm) and its coupling with H2 (4.4 Hz) assessed that the resulting structure is the desired α-furanose, instead of a pyranose. Two other molecules were detected by NMR (ca. 15% each) that corresponded to the starting Fuc and to a dibenzylated molecule (at O1 and O5).

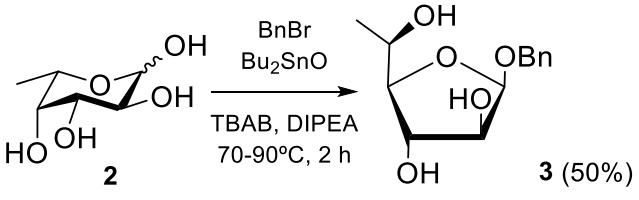


Figure 6.6. *Synthesis of intermediate 3.*

Unfortunately, only one peak was detected in the TLC using a variety of conditions, and hence the molecules could not be correctly separated. Moderate variations on the experimental conditions (temp., concentrations of

BnBr and DIPEA) did not trigger significant changes in the yield of **3**, and always similar residual amounts of both impurities were found. Alternatively, the classic methodology described by Gardiner and Percival [49-50] was then used. This strategy relies on the isolation of the different pyranose and furanose anomers of a given sugar by forming their corresponding methyl glycosides in HCl/MeOH.

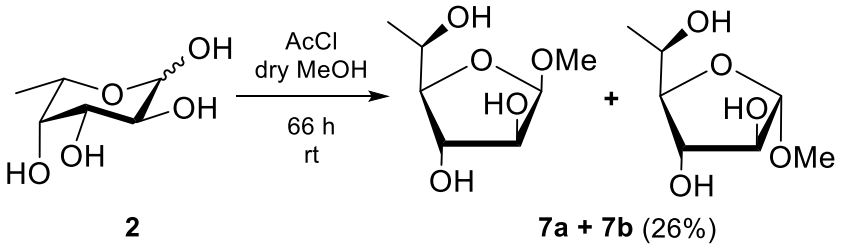


Figure 6.7. *Alternative synthesis of intermediate 3 as methoxide (7a) instead of benzyl glycoside.*

At least three products could be distinguished when accomplishing this reaction [50]. Only the most polar product detected by TLC was a pyranose (α anomer), while the other ones were fittingly the two furanoside anomers. However, the isolation of the desired methyl α-fucofuranoside entailed complications due to the very similar polarity of the three obtained compounds. The described conditions for the chromatography (EtOAc/MeOH 95:5) exclusively allowed the isolation of the first eluted compound, the β-fucofuranoside, whereas the remaining α-fucofuranoside and the pyranoside were always accompanied with notable amounts of the previous product. No other solvents or EtOAc/MeOH ratios permitted to improve this without significantly reducing the starting amount of reactant. Eventually, as the required product eluted at an intermediate R_f between the other two, the mixture of fucofuranoses was isolated (α/β 20:80) and used for the subsequent reactions, only discarding the six-membered product (Figure 6.7). In principle, both furanose anomers were valid for the required transformations [43].

The second step involved the direct formation of the epoxide at positions 2 and 3 of the fucofuranose scaffold, by using a Mitsunobu-type reaction (Figure 6.8A). Nevertheless, new complications precluded the generation of the desired epoxide either starting from **3** or **7a**. In the first case, the application of the specified conditions led to the generation of complex mixtures of products, whose separation was unworkable at all. Four to five fucose-related species were at least present in the major spot isolated from TLC, as assessed

by NMR (Figure 6.8B). In the second case, the starting mixture of methoxyfuranoses better conducted to only two major products, which presumably corresponded to compounds derived from the β-furanoside given the extremely short coupling constants displayed by both anomeric protons. The desired α-2,3-epoxifuranoside **5b** was also detected and identified by using H-1 as label (δ 4.98 ppm) [45], although it appeared as a minor byproduct (ca. 10% yield). To note, the HRMS analysis showed that the expected loss of an oxygen atom as a result of the epoxide formation never took place in high yield. The chemical nature of the starting substrate could be the reason for this, since that procedure has been described for the α-D-Ara furanoside analogue.

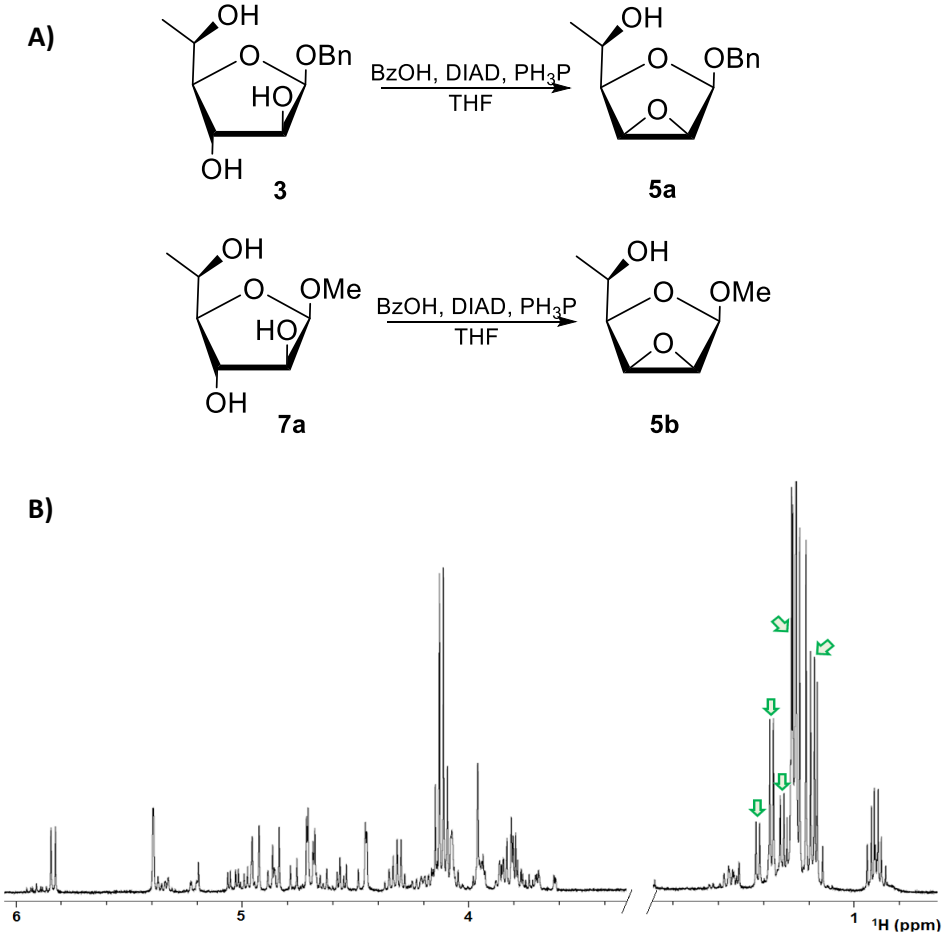


Figure 6.8. *A) Experimental procedures accomplished to try affording the intermediate 2,3-epoxifucofuranose. B)* ^{1}*H-1D spectrum recorded for the main component of the Mitsunobou reaction isolated along the flash chromatography. As noted, it consists of a complex mixture of multiple products that could not be assigned. The green arrows*

indicate some identified doublets (^3J = 5-8 Hz) corresponding to independent methyl groups (H6) from different reaction products.

The low yields, especially regarding the second step, and the presence of other undesired byproducts in large amounts led to definitely abandon this synthetic plan and seek for a more efficient one.

6.3. The alternative route from L-galactopyranose

In view of the unsuccessful results obtained from the classic methodologies, a completely new strategy was then outlined: synthesizing the 3-deoxy-3-fluoro-L-fucose through the structural modification of a different previously fluorinated sugar, more accessible in terms of time and consumables.

In particular, a detailed inspection of the stereochemistry of 3F-L-Fuc in its linear form allowed to designing a novel retrosynthetic route starting from 4F-D-Gal (Figure 6.9). Certainly, the conversion of the 4F-D-Gal template into the 3F-L-Fuc scaffold a priori involves only two modifications: the complete reduction of the anomeric carbon to yield the C6-methyl group and the oxidation of the OH6 of 4F-D-Gal, which would turn into the new anomeric center C1. The remaining question concerns the production of 4F-D-Gal, whose synthesis is actually rather simple starting from the relatively cheap methyl α-D-glucoside and using DAST as fluorinating agent [51]. As reviewed by P. J. Card [52], this synthesis only requires the previous 6-O-trityl protection [53]. The 4-deoxy-4-fluoro-D-Gal is then transformed into the corresponding cyclic thioacetal [54-55], and immediately reduced by Raney Nickel [56-58] to yield the linear 3F-fucitol. The last stage involves the oxidation of C6 of the D-Gal scaffold (now C1) to generate the carbonyl group. From a general perspective, the final oxidation would be actually the most challenging step and the major source of potential complications: the multiple hydroxyl groups are presumably susceptible of undertaking oxidation processes (selectivity) and undesired epimerization events at C2 in the linear form arising from the basic/acidic conditions employed.

Theoretically, the over-oxidation and epimerization problems could be reasonably overcome by working under mild and neutral oxidizing conditions (e.g. TEMPO or Dess-Martin). Even so, chemo-selectivity had to be carefully evaluated, since the controlled oxidation at the primary position in the presence of the many other hydroxyl groups strongly depends on the nature

of the substrate and the experimental conditions [59-61]. Overall, the synthetic plan involves few transformations (Figure 6.10).

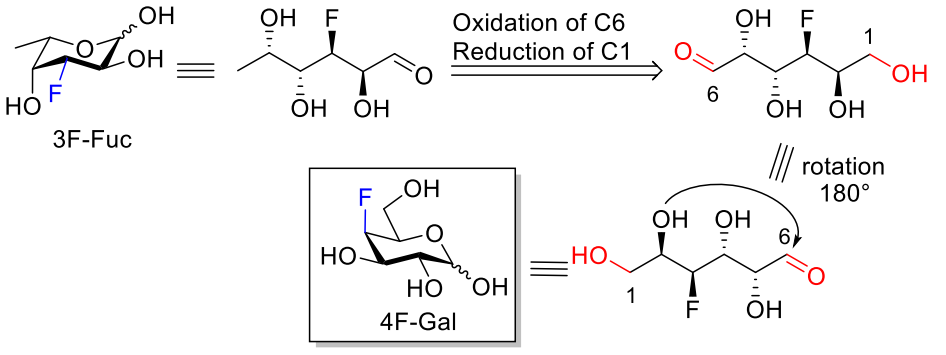


Figure 6.9. *Retrosynthetic analysis for the synthesis of 3-fluorofucose through the inversion of the terminal positions (C1 and C6) in the linear scaffold of the 4-fluorogalactose.*

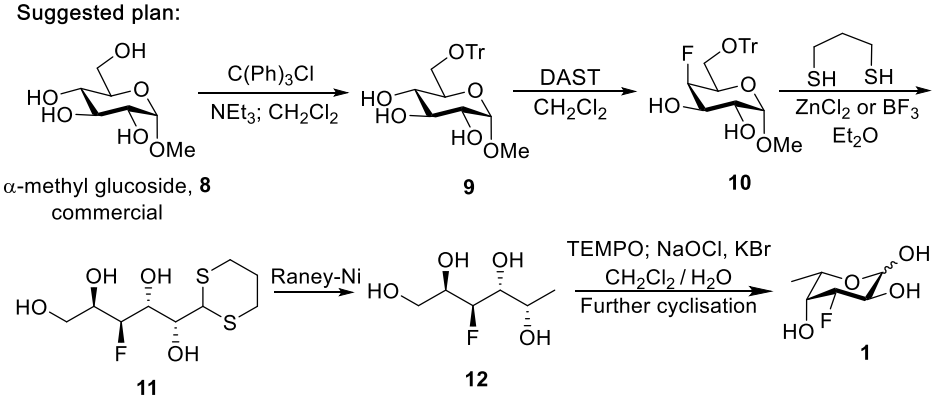
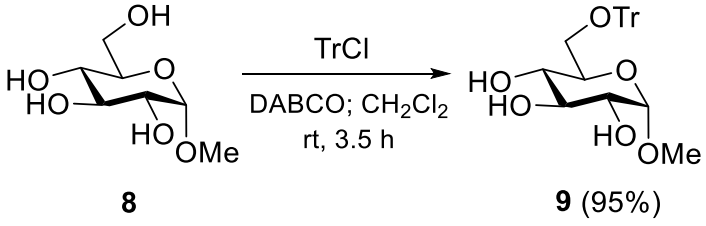


Figure 6.10. *Preliminary synthetic plan to afford the desired sugar, based on the retrosynthesis shown above. The possible protecting steps after the formation of the cyclic thioacetal are not included here.*

6.3.1. Results

The initial protection of the primary 6-OH group was straightforward, giving good yields (80-95%) of 6-OTr-α-OMe-D-glucopyranose (**9**) and no side byproducts. Worth noting, the standard conditions required to carry out this protection step (Figure 6.10) have been reported to be considerably inefficient for this particular substrate, specifically regarding the employed base, NEt_3. Hence [53], DABCO was chosen as non-nucleophilic base, and the equivalents

of the reacting trityl chloride (TrCl) and the cited base were suitably adjusted to achieve optimal yields in reasonable times (<3.5 h) (Figure 6.11).

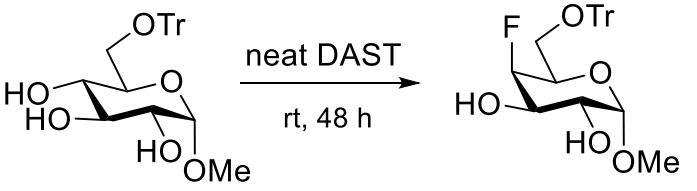


8 **9** (95%)

Figure 6.11. *Synthesis of intermediate 9.*

The presence of the trityl group was confirmed by HRMS and the proton chemical shifts already described for **9** [53] allowed verifying the nature of the obtained product. For the fluorination step, the regioselectivity was thoroughly controlled by reproducing the experimental conditions previously described [51-52]. In this regard, it is important to highlight how the regioselectivity radically changes depending on multiple factors, namely the existing protecting groups, the type of sugar, the specific anomer, and the amount of fluorinating agent. Thus, the choice of the starting substrate was not trivial: the generation of a D-Gal scaffold fluorinated at C4 firstly demanded the utilization of D-Glc as starting compound, since the reaction entails an inversion of configuration at the affected C4 carbon. Additionally, the selective entrance of a fluorine atom at C4 required the previous blockage of the primary OH at C6, and very importantly, the exclusive utilization of the α-anomer (Table on Figure 6.12). As depicted below, the major drawback of this process was obviously its low yield, whereby higher amounts of the starting material would be needed to scale up the global synthesis. Interestingly, when the same reaction was performed in neat DAST (not described), it successfully conducted to the same product and the yields slightly improved up to 30-40%.

The major product was checked to be monofluorinated by ^{19}F-NMR, and the ^{19}F chemical shift fitted with the reported value [52]. Analogously, the ^{1}H-NMR spectrum showed slight changes on the chemical shifts of the sugar protons, except for H4, which displayed a remarkable displacement towards lower fields, as well as a large geminal coupling with the adjacent fluorine (^{2}J = 50 Hz), thereby confirming the identity of the desired product.



9 **10** (38%)

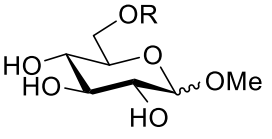
SUBSTRATE	R	Conditions	C1	Max. μ	F pos.
	R = H	Neat DAST	α	60%	4 & 6
		DAST in DCM	α	88%	6
		DAST in DCM	β	51%	6
	R = OTr	DAST in DCM	**α**	**23%**	**4**
		DAST in DCM	β	50%	3
	R = OPiv	DAST in DCM	α	-	-
		DAST in DCM	β	35%	3

Figure 6.12. *On the top, the synthesis of methyl 4-deoxy-4-fluoro-6-O-trityl-α-D-galactopyranoside performed in neat DAST. On the bottom, the different DAST-driven fluorinations that have been described [52] using methyl D-glucopyranoside as starting scaffold. The only conditions to selectively introduce the fluorine atom at C4 is highlighted in light blue.*

At this point, the subsequent steps involved the two transformations to obtain the L-Fuc scaffold, already fluorinated. The initial step consists of reacting the sugar in its linear form with 1,3-propanedithiol (PDT) to afford the corresponding cyclic dithiane at C1, which can be readily hydrogenated with Raney nickel. In this regard, it was assumed that the anomeric carbon should be free (1-OH) to enable the transient formation of the carbonyl group, which would eventually form the cyclic dithiane. For this purpose, the protected 4F-Gal was peracetylated under strong acidic conditions (including C6) and then, the acetyl moiety at C1 was selectively removed in methanolic ammonia, providing the corresponding 1-OH free peracetylated fluorosugar in moderate yields (Figure 6.13A). However, the further reaction in the presence of PDT was unsuccessful.

In parallel, when the same reaction was directly attempted on the starting methyl galactoside, the linear structure with the anomeric carbon completely blocked as a dithiane was actually obtained. According to previous reports, the conditions employed in this reaction have been demonstrated to work for methyl D-altropyranosides and for the corresponding free sugars [62]. It has

been herein proven that they actually work for other methyl glycosides, leading to the linear substrate in a unique step. Furthermore, the acidic media in which the reaction takes place expectedly triggered the loss of the trityl group at C6, which massively precipitated as trityl alcohol (TrOH). Hence, after several slight corrections in the experimental protocol (removal of TrOH), compound **11** could be readily obtained in one-pot process (protection of C1 + deprotection of C6) with good yields (Figure 6.13B).

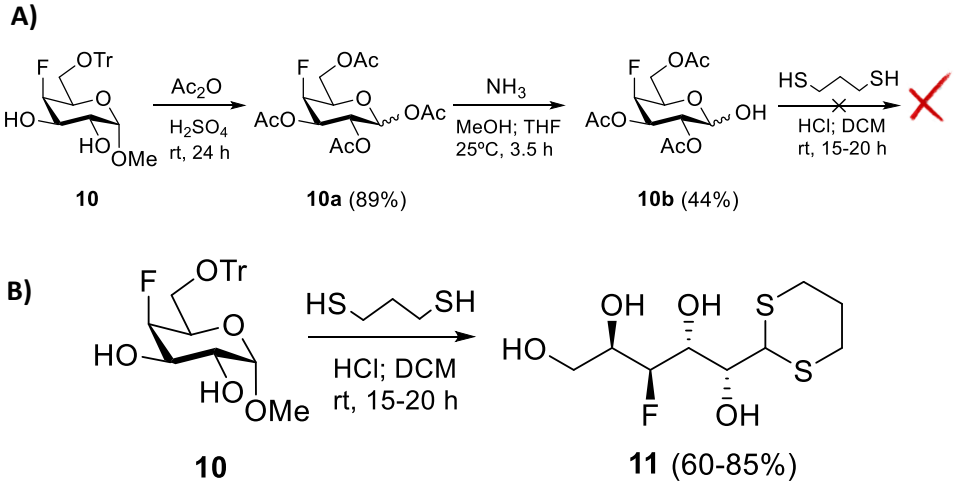


Figure 6.13. *A) Failed route to afford the required thioacetal through the free sugar. B) Successful synthetic route to obtain intermediate 11.*

Once completed this step, the reduction process was directly carried out on intermediate **11**. Intriguingly, the substrate did not efficiently evolve under the imposed conditions even after very long times, as revealed by TLC. Also, the reaction was neither well promoted in the simultaneous presence of gas hydrogen and Raney Nickel [56], and no products different from the initial compound were detected either by TLC or by NMR. Intriguingly, the ^{1}H-1D NMR spectrum displayed a striking intense and uniform band which spread over more than 1 ppm (3.2-4.6 ppm) at the region of the sugar protons (Figure 6.14). This band suggested the existence of polymeric species that could arise from the nickel-catalyzed coupling of the substrate molecules. On this basis, the starting compound was thought to be self-reacting instead of being reduced, a process that could be fostered by the presence of the many free hydroxyl groups acting as Lewis bases. To avoid these cross-reactivity problems, an intermediate protection step was included, in which the linear scaffold was transformed into the corresponding diacetonide (Figure 6.15), following standard procedures already described for analogous structures [63-

65]. Thus, the subsequent reduction was perfectly accomplished in the sole presence of Raney nickel, providing a unique and fairly clean product [66]. By NMR, the resulting product was checked to remain protected (Ac groups resonating at 1.3-1.5 ppm). As mentioned, the purity was notably high since no other fluorinated byproducts could be detected by ^{19}F-NMR. The identity of the product could be assessed both by the disappearance of the NMR peaks from the six-member thioacetal ring and the drastic change in the chemical shift of H1, which moved to the aliphatic region of the spectrum (doublet at 1.3 ppm).

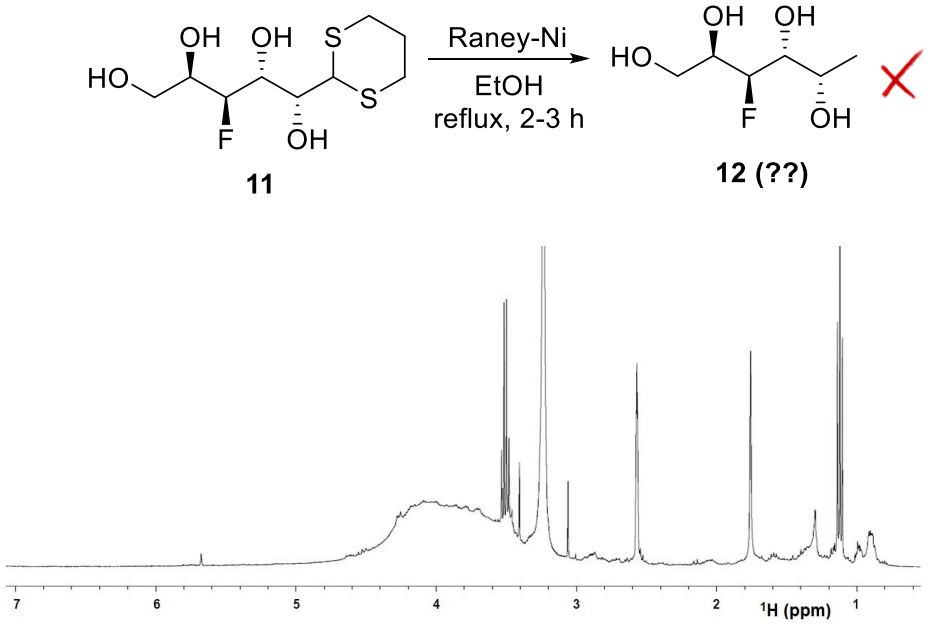


Figure 6.14. *On the top, the reduction process performed on the unprotected polyol. On the bottom, the ^{1}H-1D NMR of the reaction crude isolated after the reaction. As observed, there is an extremely broad band at the central area, between 3 and 4 ppm.*

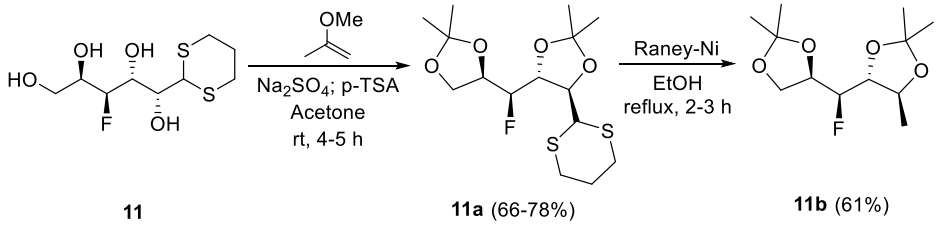


Figure 6.15. *Two-step reduction of C1 from thioacetal to methyl group.*

Before the final oxidation of the opposite carbon (C6), the intermediate **11b** was easily hydrolyzed into the pure 3-deoxy-3-fluoro-L-fucitol **12** [67-68]

(Figure 6.16). Different options were considered to selectively oxidate the primary position of the unprotected substrate under mild conditions [69-70], including not only TEMPO-mediated processes [71-73], but also other alternatives as the Dess-Martin periodinane (DMP) [74]. Unfortunately, none of these options gave positive results. The oxidation mediated by TEMPO/NaOCl [72] led in most of cases to the recovery of the starting fucitol and to a mixture of unidentified compounds. Analogously, although the reaction progressed in the presence of either TEMPO/TCCA [71] or DMP [74] and the final sugar (**1**) could be actually detected, it only appeared as a minor compound in a complex mixture (Figure 6.17).

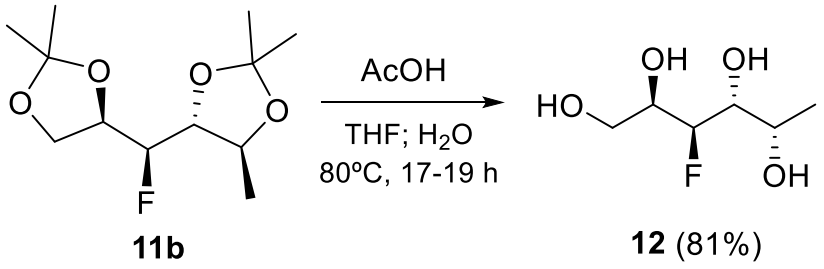


Figure 6.16. *Full deprotection of diacetonide* ***11b.***

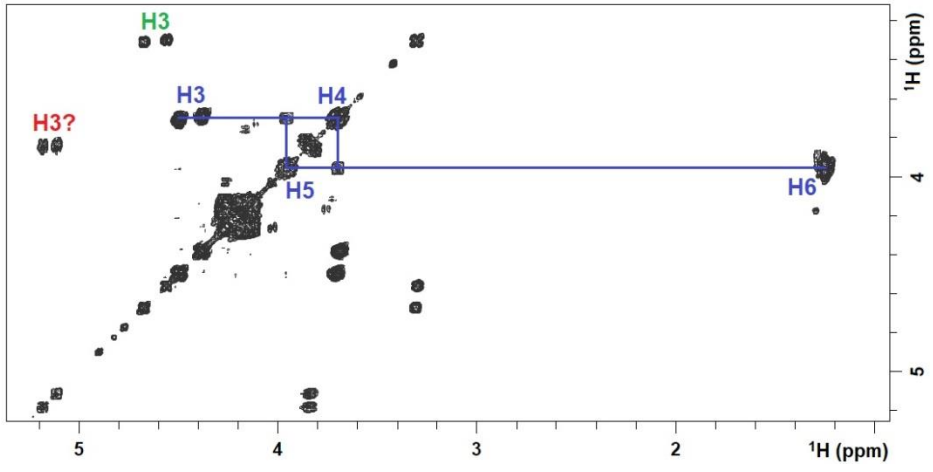


Figure 6.17. *Major products isolated from the treatment of compound* ***12*** *with DMP (in pyridine and DCM, 0ºC, 3-4 hours). The* ^{19}*F-{*^{1}*H}-NMR showed a major peak at -208.6 ppm (blue protons). Both the* ^{13}*C-{*^{1}*H}-NMR and the 2D-COSY suggested the loss of a carbon atom, later confirmed by HRMS. The resulting product was probably the carboxylic acid (peak at 171 ppm from a carbonyl group) obtained by oxidative cleavage of the C1-C2 bond, as supported by the lack of correlations from the fluorinated carbon.*

As a potential solution to reduce the possible over-oxidation problems, partial deprotection of diacetonide **11b** was attempted [75]. However, the classic acidic hydrolysis using AcOH did not result effective [76-77], since it led to the complete deprotection of the diacetonide intermediate, providing compound **12** as well. Other alternatives based on catalysis with a Lewis acid ($CuCl_2$) [78] utterly failed and yielded the starting substrate.

At this point, it was decided to investigate an enzymatic oxidation protocol instead of carrying on the quest for a different protecting group strategy. In this regard, the Galactose Oxidase (GOase) enzyme was evaluated as a feasible alternative. While GOase itself is strictly limited to oxidizing the C6 hydroxyl of non-reducing galactosides and a small number of aromatic primary alcohols, previous engineering efforts have greatly expanded its substrate scope to a large number of primary and secondary alcohols [79-80]. To note, this enzymatic alcohol oxidation protocol generates H_2O_2, whereby the activity of several GOase variants could be tested by monitoring the H_2O_2 production through a colorimetric assay using Horse Radish Peroxidase (HRP) [81]. This same enzyme was also needed for GOase activation. After several tests, the F_2 variant [82] was noticed to show a high activity (100 fold higher than other tested variants), successfully driving the conversion of 3-deoxy-3-fluoro-L-fucitol into **3F-Fuc (1)** in good yield on mmol scale (Figure 6.18), in the presence of HRP and catalase. In view of these results, the process can be further scaled up.

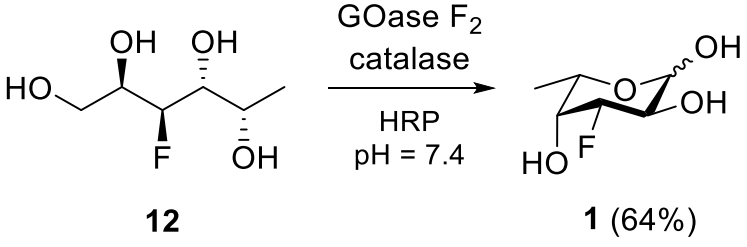


Figure 6.18. *Enzymatic synthesis of 3F-Fuc directly from 3F-L-fucitol* ***12****.*

Compound **1** was found to exist as a 1:2 mixture of anomers by ^{19}F-NMR in favor of the β-anomer (in D_2O), which is essentially the same ratio observed for non-fluorinated Fuc [83]. A basic analysis of the coupling constants (Figure 6.19) indicated that both anomers exist in the expected $^1C_4$ conformation: the $^3J_{H1\text{-}H2}$ value was large for the β-anomer and small for the α-anomer, and both showed a large $^3J_{H2\text{-}H3}$ (axial H2 and H3). The equatorial position of the fluorine was confirmed by the relatively small $^3J_{F\text{-}H4}$ value.

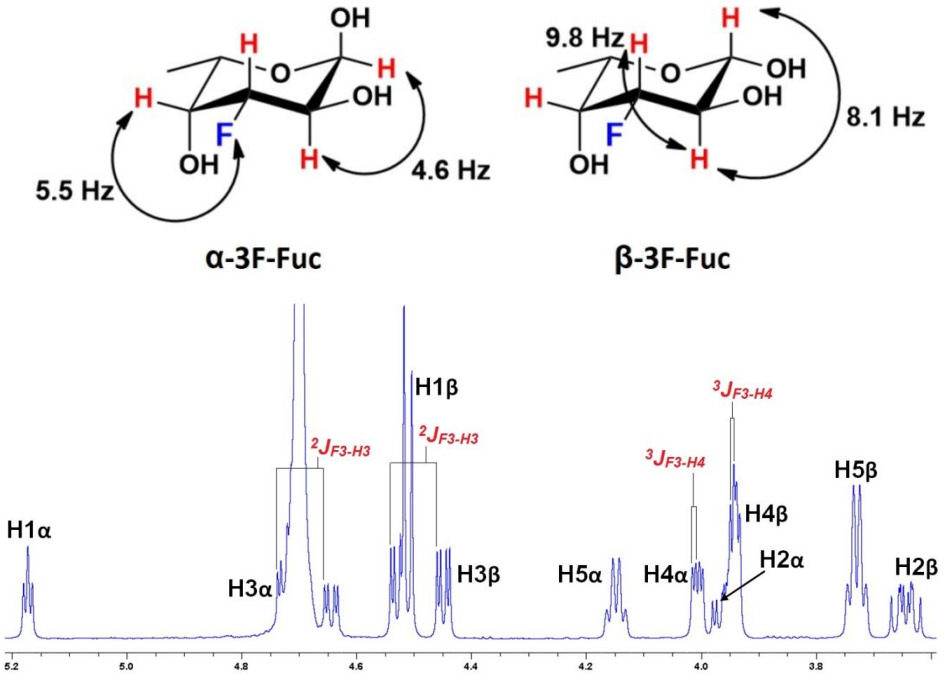
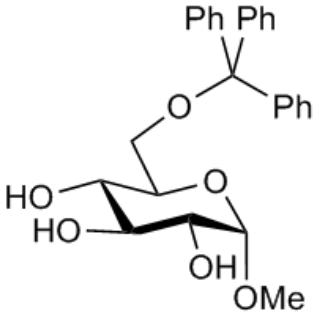


Figure 6.19. *^{1}H-NMR analysis of the final 3F-Fuc in D_2O. Full assignment in Figure 6.20.*

In summary, the first synthesis of 3-deoxy-3-fluoro-L-fucose has been achieved by using a D- to L-sugar conversion strategy, in 7 steps starting from a cheap D-glucose starting material. This strategy allowed dispensing with the common fluorination scheme based in an alcohol inversion step prior to the deoxyfluorination operation.

6.4. Experimental part

- Synthesis of **compound 9**:



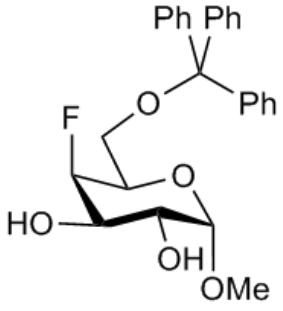
The starting methyl D-glucopyranoside (20 g, 103 mmol) was firstly mixed with DABCO (2 eq) in 400 mL of DCM, and stirred for 20-30 min under inert atmosphere at 30 ºC until it became a homogeneous suspension containing both solid components. Then, the TrCl (2 eq) was added and the reaction was allowed to progress overnight (~12-16 hours). The final product was isolated as a pure white solid by flash chromatography using Acetone/Pet. Ether 6:4 as eluent (R_f = 0.32).

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.47-7.45 (m, 5H, Ar), 7.32-7.24 (m, 10H, Ar), 4.77 (d, J = 3.8 Hz, 1 H, H_1), 3.70-3.64 (m, 2H, H_3 and H_5), 3.55-3.50 (m, 2H, H_2 and H_4), 3.43 (s, 3H, OCH_3), 3.41 (dd, J = 10.1, 4.3 Hz, 1H, H_{6a}), 3.37 (dd, J = 10.1, 5.0 Hz, 1H, H_{6b}), 2.56 (d, J = 2.6 Hz, 1H, OH) ppm.

^{13}C NMR (125 MHz, $CDCl_3$): δ = 143.7, 128.6, 128.0, 127.9, 127.9, 127.2 (Ph), 99.1 (C1), 87.1 [$\underline{C}(Ph)_3$], 74.8 (C3), 72.3 (C2), 71.9 (C4), 69.8 (C5), 64.0 (C6), 55.3 ($O\underline{C}H_3$) ppm.

HRMS (ESI+) (m/z): [M+Na^+] calcd. for $C_{26}H_{28}NaO_6$, 459.1784; found 459.4593.

- Synthesis of **compound 10**:



Compound **9** (25 g, 57.3 mmol) was slowly added to a flask containing 69 mL of DAST (9 eq) while cooling down the system below 0 ºC. The reaction flask was sealed and the reaction was allowed to progress under argon atmosphere for at least 48 hours (until no further significant changes were noticed by TLC). Once finished, the reaction crude was quenched by firstly diluting the remaining DAST with DCM (140 mL), and subsequently neutralizing it with MeOH (60-70 mL). Then, water was also added and the crude was firstly extracted with sat. $NaHCO_3$, neutralized with HCl and washed. The resulting syrup was eventually purified by flash chromatography (DCM/Acetone 85:15), which yielded a white solid as the final product (R_f = 0.40)

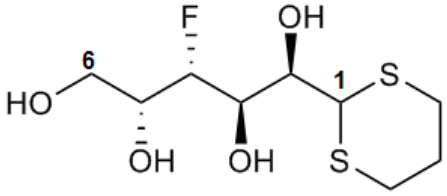
^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.45-7.43 (m, 5H, Ar), 7.32-7.22 (m, 10H, Ar), 4.92 (dd, J = 50.4, 2.3 Hz, 1H, H_4), 4.80 (d, J = 3.2 Hz, 1H, H_1), 3.81-3.74 (m, 2H, H_3 and H_2), 3.75 (dt, J = 30.2, 6.6 Hz, 1H, H_5), 3.41 (s, 3H, OCH_3), 3.39 (ddd, J = 9.5, 6.2, 0.9 Hz, 1H, H_{6a}), 3.36 (dd, J = 9.4, 7.0 Hz, 1H, H_{6b}), 2.28 (dd, J = 6.0, 1.0 Hz, 1H, OH), 2.02 (d, J = 9.0 Hz, 1H, OH) ppm.

^{13}C NMR (125 MHz, $CDCl_3$): δ = 143.8, 128.7, 128.0, 127.3 (Ph), 99.3 (C1), 90.1, 88.6 (J = 181.3 Hz, C2), 87.2 [$\underline{C}(Ph)_3$], 70.5, 70.3 (J = 17.5 Hz, C3), 70.1, 70.0 (J = 2.2 Hz, C2), 68.8, 68.6 (J = 18.6 Hz, C5), 61.7 (C6), 55.7 ($O\underline{C}H_3$) ppm.

^{19}F NMR (472 MHz, $CDCl_3$): δ = -221.1 (F_4) ppm.

HRMS (ESI+) (m/z): [M+Na^+] calcd. for $C_{26}H_{27}FNaO_5$, 461.1740; found 461.3918.

- Synthesis of **compound 11**:



Compound **10** (6.31 g, 14.4 mmol) was dissolved in 20 mL of CH_3Cl and 20 mL of conc. HCl. Under vigorous stirring, 2.5 mL of PDT (1.5 eq) were dropwise added and the reaction was allowed to progress at room temperature for 20 hours. Afterwards, the crude was neutralized by adding saturated $NaHCO_3$ and cooled down to promote the precipitation of the generated TrOH. The cited byproduct was filtered off and the crude was dried out and re-suspended in DCM before the flash column. The desired thioacetal was isolated as yellowish oil by using DCM/MeOH 95:5 as eluent (R_f = 0.28).

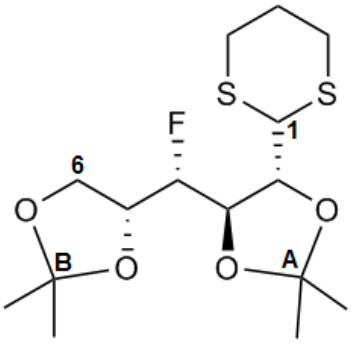
^{1}H NMR (500 MHz, MeOD): δ = 4.62 (ddd, J = 46.1, 9.5, 1.2 Hz, 1H, H_4), 4.40 (ddd, J = 9.5, 3.7, 0.9 Hz, 1H, H_3), 4.15 (d, J = 9.8 Hz, 1H, H_1), 3.99-3.90 (m, 2H, H_2 and H_5), 3.66 (d, J = 6.8 Hz, 2H, H_6), 2.97-2.87 (m, 2H, CH_2-S), 2.78-2.72 (m, 2H, CH_2-S), 2.09-2.02 (m, 1H, CH_2-C), 1.96-1.88 (m, 1H, CH_2-C) ppm.

^{13}C NMR (125 MHz, MeOD): δ = 92.1, 90.7 (J = 176.3 Hz, C4), 70.8, 70.7 (J = 18.2 Hz, C5), 70.7, 70.7 (J = 2.7 Hz, C2), 67.8, 67.6 (J = 26.3 Hz, C3), 63.6, 63.5 (J = 5.7 Hz, C6), 49.0 (C1), 28.8 (CH_2-S), 28.5 (CH_2-S), 27.0 ($\underline{C}H_2$-C) ppm.

^{19}F NMR (472 MHz, MeOD): δ = -213.3 (F_4) ppm.

HRMS (ESI+) (m/z): [M^+] calcd. for $C_9H_{17}FO_4S_2$, 272.0552; found 272.2437.

- Synthesis of **compound 11a**:



The starting molecule **11** (2.36 g, 8.67 mmol), Na_2SO_4 (cat.) and p-TsOH (cat.) were directly dissolved in 35 mL of acetone, and finally 2 mL of 2-methoxypropene (2 eq) were added before sealing the flask. The working mixture was then stirred overnight (16-20 hours) at room temperature under argon atmosphere, until the reaction was completed. The resulting crude was firstly quenched with solid K_2CO_3, filtered and resuspended in Pet. Ether/EtOAc 9:1. Using the same system as eluent, the final diacetonide was isolated by flash chromatography as yellow oil (R_f = 0.24).

^{1}H NMR (500 MHz, $CDCl_3$): δ = 4.46 (dd, J = 6.1, 3.7 Hz, 1H, H_2), 4.44-4.32 (m, 1H, H_4*), 4.36-4.31 (m, 1H, H_3), 4.31 (dtd, J = 24.2, 6.8, 3.8 Hz, 1H, H_5), 4.15 (d, J = 3.6 Hz, 1H, H_1), 4.10 (td, J = 8.5, 6.8 Hz, 1H, H_{6a}), 3.96 (dd, J = 8.5, 6.8 Hz, 1H, H_{6b}), 3.05-2.99 (m, 2H, CH_2-S), 2.83-2.73 (m, 2H, CH_2-S), 2.12-2.05 (m, 1H, CH_2-C), 2.03-1.95 (m, 1H, CH_2-C), 1.49 (d, J = 0.6 Hz, 3H, Me_A), 1.43 (s, 3H, Me_B), 1.39 (d, J = 0.6 Hz, 3H, Me_A), 1.39 (d, J = 0.6 Hz, 3H, Me_B) ppm.

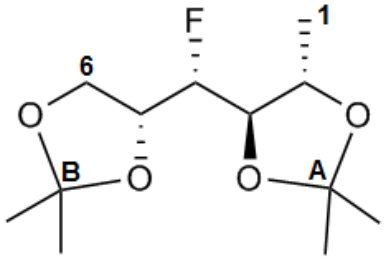
^{13}C NMR (125 MHz, $CDCl_3$): δ = 111.4 (C_A), 110.1 (C_B), 93.3, 91.8 (J = 181.6 Hz, C4), 83.1 (C2), 76.6, 76.4 (J = 27.8 Hz, C3), 75.2, 75.1 (J = 16.9 Hz, C5), 65.2, 65.1 (J = 7.0 Hz, C6), 47.7 (C1), 29.6 (CH_2-S), 29.1 (CH_2-S), 27.4 (Me_A), 27.4 (Me_A), 26.2 (Me_B), 25.9 ($\underline{C}H_2$-C), 25.8 (Me_B) ppm.

^{19}F NMR (472 MHz, $CDCl_3$): δ = -204.5 (F_4) ppm.

*The $^2J_{H4\text{-}F4}$ could be determined in the ^{13}C-HSQC: 48.3 Hz.

HRMS (ESI+) (m/z): [M+H$^+$] calcd. for $C_{15}H_{25}FO_4S_2$, 353.1279; found 353.3824.

- Synthesis of **compound 11b**:



Compound **11a** (5.40 g, 15.34 mmol) was firstly dissolved in a suspension of Raney Ni in EtOH (135 mL). The starting mixture was diluted with additional 210 mL of EtOH and heated at reflux (80 ºC) under continuous stirring. The reaction was checked to be moderately completed after 2-3 hours, whereby the reaction times were further raised to 16-24 hours to maximize the yields. The obtained crude was then filtered through 2-3 cm of celite to retain the Raney Ni, and washed in the same filter with DCM, EtOH (three times) and finally DCM/MeOH (1:1). The reduced compound **11b** could be straightforwardly purified by flash chromatography, using DCM/MeOH 95:5 as eluent (R_f = 0.32). The product was a colorless liquid with high purity.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 4.36-4.24 (m, 2H, H_4* and H_5*), 4.17 (qd, J = 6.0, 0.8 Hz, 1H, H_2), 4.09 (ddd, J = 8.5, 6.7, 1.3 Hz, 1H, H_{6a}), 3.94 (dd, J = 8.5, 6.7 Hz, 1H, H_{6b}), 3.73 (td, J = 7.6, 5.6 Hz, 1H, H_3), 1.43 (s, 3H, Me_B), 1.42 (d, J = 0.5 Hz, 3H, Me_A), 1.39 (d, J = 0.5 Hz, 3H, Me_B), 1.38 (dd, J = 6.1, 0.9 Hz, 3H, H_1), 1.36 (d, J = 0.5 Hz, 3H, Me_A) ppm.

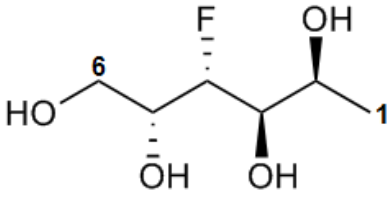
^{13}C NMR (125 MHz, $CDCl_3$): δ = 109.9 (C_B), 109.3 (C_A), 93.6, 92.1 (J = 181.5 Hz, C4), 79.4, 79.2 (J = 31.3 Hz, C3), 76.2 (C2), 75.3, 75.2 (J = 17.8 Hz, C5), 65.1, 65.0 (J = 6.6 Hz, C6), 27.4 (Me_A), 26.7 (Me_A), 26.1 (Me_B), 25.6 (Me_B), 18.8, 18.8 (J = 2.2 Hz, C1) ppm.

^{19}F NMR (472 MHz, $CDCl_3$): δ = -205.4 (F_4) ppm.

* $^{2}J_{H4\text{-}F4}$ = 46.1 Hz; $^{2}J_{H5\text{-}F4}$ = 24.7 Hz. Determined in the ^{13}C-HSQC.

HRMS (ESI+) (m/z): [M+Na^+] calcd. for $C_6H_{13}FNaO_4$, 271.1322; found 271.1690.

- Synthesis of **compound 12**:



The diacetonide **11b** (0.51 g, 2.06 mmol) was dissolved in 12.2 mL of a THF/H_2O 1:1 solution. Under gentle stirring, 12.1 mL of glacial AcOH (> 50 eq) were added and the resulting mixture was then heated overnight at reflux (80 ºC). The final crude could be readily concentrated and columned on silica. DCM/MeOH 9:1 was used to elute compound **12** (R_f = 0.26) with high purity as a white solid.

^{1}H NMR (500 MHz, MeOD): δ = 4.54 (ddd, J = 46.1, 9.1, 1.1 Hz, 1H, H_4), 4.00-3.89 (m, 2H, H_2 and H_5), 3.67 (dd, J = 4.5, 1.9 Hz, 1H, H_3), 3.65 (d, J = 6.8 Hz, 2H, H_6), 1.25 (d, J = 6.7 Hz, 3H, H_1) ppm.

^{13}C NMR (125 MHz, MeOD): δ = 92.4, 91.0 (J = 175.7 Hz, C4), 72.5, 72.3 (J = 25.7 Hz, C3), 71.3, 71.2 (J = 18.1 Hz, C5), 67.0, 67.0 (J = 2.6 Hz, C2), 63.7, 63.6 (J = 5.7 Hz, C6), 20.0 (C1) ppm.

^{19}F NMR (472 MHz, MeOD): δ = -212.7 (F_4) ppm.

HRMS (ESI+) (m/z): [M+Na^+] calcd. for $C_6H_{13}FNaO_4$, 191.0695; found 191.0690.

- Characterization of **compound 1**:

^{1}H NMR (600 MHz, D_2O): δ = 5.17 (t, J = 4.6 Hz, 1H, $H_1\alpha$), 4.69 (ddd, J = 49.2, 10.5, 3.1 Hz, 1H, $H_3\alpha$), 4.51 (d, J = 8.1 Hz, 1H, $H_1\beta$), 4.49 (ddd, J = 48.3, 9.7, 3.6 Hz, 1H, $H_3\beta$), 4.15 (q, J = 6.7 Hz, 1H, $H_5\alpha$), 4.01 (dd, J = 7.3, 3.5 Hz, 1H, $H_4\alpha$), 3.95 (ddd, J = 14.4, 10.3, 4.1 Hz, 1H, $H_2\alpha$), 3.94 (dd, J = 6.3, 3.5 Hz, 1H, $H_4\beta$), 3.73 (t, J = 6.5 Hz, 1H, $H_5\beta$), 3.64 (ddd, J = 8.1 Hz, 1H, $H_2\beta$), 1.19 (d, J = 6.5 Hz, 3H, $H_6\beta$), 1.15 (d, J = 6.7 Hz, 3H, $H_6\alpha$) ppm.

^{19}F NMR (498 MHz, D_2O): δ = -198.5 ($F_3\beta$), -202.3 ($F_3\alpha$) ppm.

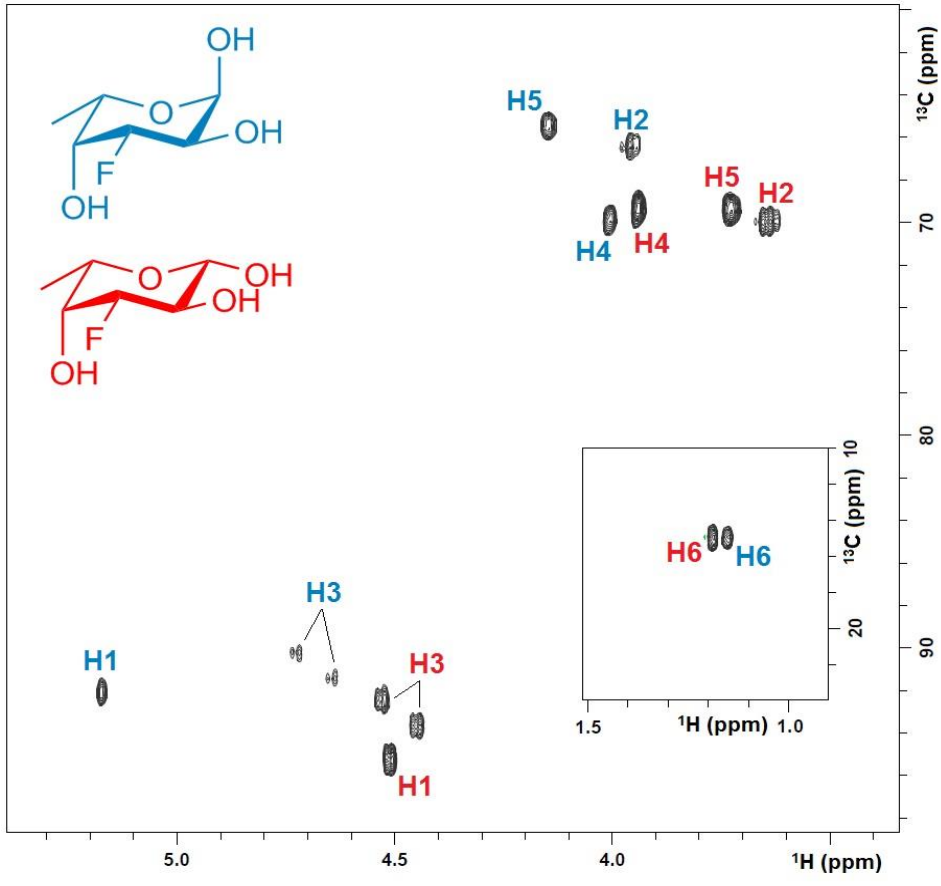


Figure 6.20. *^{13}C-HSQC full assignment of both anomers of compound* ***1.***

6.5. References

1. a) P. J. Card. Synthesis of fluorinated carbohydrates. *J. Carbohydr. Chem.,* (1985), **4**, 451–487. b) T. Tsuchiya. Chemistry and developments of fluorinated carbohydrates. *Adv. Carbohydr. Chem. Biochem.,* (1990), **48**, 91–277. c) K. Dax, M. Albert, J. Ortner and B. J. Paul. Synthesis of deoxyfluoro sugars from carbohydrate precursors. *Carbohydr. Res.,* (2000), **327**, 47-86.
2. a) M. D. Burkart, Z. Zhang, S.-C. Hung and C.-H. Wong. A new method for the synthesis of fluoro-carbohydrates and glycosides using Selectfluor. *J. Am. Chem. Soc.,* (1997), **119**, 11743–11746. b) S. Cai, S. Hakomori and T. Toyokuni. Application of protease-catalyzed regioselective esterification in synthesis of 6'-deoxy-6'-fluoro- and 6-deoxy-6-fluorolactosides. *J. Org. Chem.,* (1992), **57**, 3431–3437.
3. L. A. Mulard, P. Kovac and C. P. J. Glaudemans. Synthesis of specifically monofluorinated ligands related to the O-polysaccharide of *Shigella dysenteriae* type 1. *Carbohydr. Res.,* (1994), **259**, 21-34.
4. a) P. Kovác. A short synthesis of 2-deoxy-2-fluoro-D-glucose. *Carbohydr. Res.,* (1986), **153**, 168–170. b) C. Simiand and H. Driguez. Synthesis of sucrose analogues modified at position 4. *J. Carbohydr. Chem.,* (1995), **14**, 977–983.
5. a) T. Haradahira, M. Maeda, Y. Yano and M. Kojima. New synthesis of 2-deoxy-2-fluoro-D-galactose. *Chem. Pharm. Bull.,* (1984), **32**, 3317-3319. b) G. D. Daves Jr., P. Kovác and C. P. J. Glaudemans. Synthesis of the 4-deoxy-2-deutero-4-fluoro analog of methyl O-β-D-galactopyranosyl-(1→6)-β-D-galactopyranoside. *J. Carbohydr. Chem.,* (1990), **9**, 101–112. c) K. Koch and R. J. Chambers. An improved synthesis of 4-deoxy-4-fluoro-d-galactopyranosyl derivatives. *Carbohydr. Res.,* (1993), **241**, 295–299.
6. a) J. Adamson and A. B. Foster. Fluorinated carbohydrates. Part III. 2-deoxy-2-fluoro-D-glucose and 2-deoxy-2-fluoro-D-mannose. *Carbohyd. Res.,* (1970), **15**, 351-359. b) T. Haradahira, M. Maeda, H. Omae, Y. Yano and M. Kojima. Synthesis of 2-deoxy-2-fluoro-D-mannose using fluoride ion. *Chem. Pharm. Bull.,* (1984), **32,** 4758–4766.
7. a) P. Herdewijn, A. van Aerschot and L. Kerremans. Synthesis of Nucleosides Fluorinated in the Sugar Moiety. The application of diethylaminosulfur trifluoride to the synthesis of fluorinated nucleosides. *Nucleosides Nucleotides Nucleic Acids,* (2006), **8**, 65-96. b) A. Cavaliere, K. C. Probst, A. D. Westwell and M. Slusarczyk. Fluorinated nucleosides as an important class of anticancer and antiviral agents. *Future Med. Chem.*, (2017), **9**, 1809-1833.
8. D. J. Becker and J. B. Lowe. Fucose: biosynthesis and biological function in mammals. *Glycobiology,* (2003), **13**, 41R-53R.
9. S. Marionneau, A. Cailleau-Thomas, J. Rocher, B. Le Moullac-Vaidye, N. Ruvoën, M. Clément and J. Le Pendu. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie.,* (2001), **83**, 565-573.

10. M. Rinnbauer, B. Ernst, B. Wagner, J. Magnani, A. J. Benie and T. Peters. Epitope mapping of sialyl LewisX bound to E-selectin using saturation transfer difference NMR experiments. *Glycobiology*, (2003), **13**, 435-443.
11. A. Holla and A. Skerra. Comparative analysis reveals selective recognition of glycans by the dendritic cell receptors DC-SIGN and Langerin. *Protein Eng. Des. Sel.,* (2011), **24**, 659-669.
12. K. Pederson, D. A. Mitchell and J. H. Prestegard. Structural characterization of the DC-SIGN−LewisX complex. *Biochemistry,* (2014), **53**, 5700−5709.
13. H. Feinberg, M. E. Taylor, N. Razi, R. McBride, Y. A. Knirel, S. A. Graham, K. Drickamer and W. I. Weis. Structural basis for langerin recognition of diverse pathogen and mammalian glycans through a single binding site. *J. Mol. Biol.*, (2011), **405**, 1027–1039.
14. Y. Imahori, T. Ido, K. Ishiwata, T. Takahashi, K. Kawashima, K. Yanai, Y. Miura, M. Monma, S. Watanuki and A. Ujiie. 2-deoxy-2-[^{18}F]-fluoro-L-fucose, a potential agent for regional fucose utilization studies associated with glycoprotein synthesis. *CYRIC Annual Report,* (1984), 119-131.
15. C. G. Butchard and P. W. Kent. 2-deoxy-2-fluoro-L-fucose: synthesis and structure. *Tetrahedron,* (1979), **35**, 2551-2554.
16. W. Korytnyk, S. Valentekovic-Horvatah and C. R. Petrie III. A convenient synthesis of 1,2-difluoro-1, 2-dideoxyhexoses using xenon fluoride. *Tetrahedron,* (1982), **38**, 2547-2550.
17. M. Albert, B. J. Paul and K. Dax. Synthesis of (2-deoxy-2-fluoro-glycosyl)amino-acids. *Synlett.,* (1999), **9,** 1483–1485.
18. S. P. Vincent, M. D. Burkart, C.-Y. Tsai, Z. Zhang and C.-H. Wong. Electrophilic fluorination−nucleophilic addition reaction mediated by Selectfluor: mechanistic studies and new applications. *J. Org. Chem.*, (1999), **64**, 5264–5279.
19. J. R. Sufrin, R. J. Bernacki, M. J. Morin and W. Korytnyk. Halogenated L-fucose and D-galactose analogues: synthesis and metabolic effects. *J. Med. Chem.,* (1980), **23**, 143-149.
20. R. C. Bansal, B. Dean, S. Hakomori and T. Toyokuni. Synthesis of trifluoromethyl analogue of L-fucose and 6-deoxy-D-altrose. *J. Chem. Soc., Chem. Commun.,* (1991), 796-798.
21. M. D. Burkart, S. P. Vincent, A. Düffels, B. W. Murray, S. V. Ley and C. H. Wong. Chemo-enzymatic synthesis of fluorinated sugar nucleotide: useful mechanistic probes for glycosyltransferases. *Bioorg. Med. Chem.,* (2000), **8**, 1937–1946.
22. C. D. Rillahan, A. Antonopoulos, C. T. Lefort, R. Sonon, P. Azadi, K. Ley, A. Dell, S. M. Haslam and J. C. Paulson. Global metabolic inhibitors of sialyl- and fucosyltransferases remodel the glycome. *Nat. Chem. Biol.,* (2012), **8**, 661-668.
23. D. J. Winterbourne, C. G. Butchard and P. W. Kent. 2-deoxy-2-fluoro-L-fucose and its effect on L-[1-14C]-fucose utilization in mammalian cells. *Biochem. Biophys. Res. Commun.,* (1979), **87**, 989-992.
24. Y. Zhou, T. Fukuda, Q. Hang, S. Hou, T. Isaji, A. Kameyama and J. Gu. Inhibition of fucosylation by 2-fluorofucose suppresses human liver cancer HepG2 cell

proliferation and migration as well as tumor formation. *Sci. Rep.,* (2017), **7**, 11563.

25. M. A. Carrascal, M. Silva, J. S. Ramalho, C. Pen, M. Martins, C. Pascoal, C. Amaral, I. Serrano, M. J. Oliveira, R. Sackstein and P. A. Videira. Inhibition of fucosylation in human invasive ductal carcinoma reduces E-selectin ligand expression, cell proliferation, and ERK1/2 and p38 MAPK activation. *Mol. Oncol.,* (2018), **12**, 579-593.
26. M. L. Disis, L. Rastetter, E. Gad, M. Koehnlein, P. D. Senter, S. Gardai and N. M. Okeley. Modulation of immunity with 2-fluorofucose (2FF) for breast cancer treatment and prevention. *Cancer Res.,* (2016), **76**. Abstract PD3-08.
27. J. D. Belcher, C. Chen, J. Nguyen, F. Abdulla, P. Nguyen, M. Nguyen, N. M. Okeley, D. R. Benjamin, P. D. Senter and G. M. Vercellotti. The fucosylation inhibitor, 2-fluorofucose, inhibits vaso-occlusion, leukocyte-endothelium interactions and NF-κB activation in transgenic sickle mice. *PLoS One,* (2015), **10**, e0117772.
28. S. Caille and J. G. Allen. Discovery of novel protein fucosylation inhibitors and development of a manufacturing process to prepare inhibitor 6,6,6-trifluorofucose. *ACS Symposium Series,* (2016), **1239**, 37-68.
29. A. C. Richardson and J. M. Williams. Selective acylation of pyranosides-II. Benzoylation of methyl 6-deoxy-α-L-galactopyranoside and methyl 6-deoxy-α-L-mannopyranoside. *Tetrahedron,* (1967), **23**, 1641-1646.
30. P. Valverde, S. Delgado, J. D. Martínez, J.-B. Vendeville, J. Malassis, B. Linclau, N. C. Reichardt, F. J. Cañada, J. Jiménez-Barbero and A. Ardá. Molecular insights into DC-SIGN binding to self-antigens: the interaction with the blood group A/B antigens. *ACS Chem. Biol.,* (2019), **14**, 1660−1671.
31. E. Umemura, T. Tsuchiya, Y. Kobayashi and K. Tanaka. A synthetic study of methyl 3-deoxy-3-fluoro-alpha-D-glucopyranosides from methyl 2,3-anhydro-alpha-D-allopyranosides, and synthesis of 3'-deoxy-3'-fluorokanamycin A and 3'-chloro-3'-deoxykanamycin A. *Carbohydr. Res*., (1992), **224**, 141–163.
32. Y. Kobayashi, T. Tsuchiya, T. Ohgi, N. Taneichi and Y. Koyama. Study on fluorination of 2,3-dideoxy-2,3-(N-tosylepimino)-alpha-D-allopyranosides, and synthesis of 3'-deoxy-3'-fluorokanamycin B and 3',4'-dideoxy-3'-fluorokanamycin B. *Carbohydr. Res*., (1992), **230**, 89–105.
33. P. Fernández, J. Jiménez-Barbero and M. Martín-Lomas. Syntheses of all the possible monomethyl ethers and several deoxyhalo analogues of methyl beta-lactoside as ligands for the Ricinus communis lectins. *Carbohydr. Res*., (1994), **254**, 61–79.
34. I. A. Mikhailopulo and G. G. Sivets. Reactions of (diethylamino)sulfur trifluoride (DAST) with methyl-5-O-benzyl-β-D-*arabino*- and -*xylo*-furanosides. *Synlett*, (1996), **2**, 173–174.
35. P. Nebois, D. Picq and D. Anker. Syntheses of 3-deoxy-3-fluoroglucosamine derivatives. *J. Carbohydr. Chem*., (1990), **9,** 359–368.
36. M. S. Motawia and E. B. Pedersen. A short route to 3’-deoxy-3’-fluorothymidine. *Liebigs Ann. Chem*., (1990), **11**, 1137–1139.

37. G. Saischek, F. Fuchs, K. Dax and G. Billani, *Eur. Pat. Appl*. 91105231.4 (03.04.91) 0450585 A2 [CA 116: 42980].
38. K. Vanhessche, E. van der Eycken, M. Vandewalle and H. Röper. L-ribulose: A novel chiral pool compound. *Tetrahedron Lett*., (1990), **31**, 2337–2340.
39. K. Ok, Y. Takagi, T. Tsuchiya, S. U-Wa and H. Umezawa. Synthesis of antitumor-active 7-O-(2, 6-dideoxy-2-fluoro-α-L-talopyranosyl)-daunomycinone and -adriamycinone. *Carbohydr. Res.,* (1987), **169**, 69-81.
40. N. Yan, Z. Lei, J. Su, W. Liao and X. Hua. Ring opening of sugar-derived epoxides by TBAF/KHF2: An attractive alternative for the introduction of fluorine into the carbohydrate scaffold. *Chin. Chem. Lett.*, (2017), **28***,* 467–470.
41. M. H. Lim, H. O. Kim, H. R. Moon, S. J. Lee, M. W. Chun, Z.Gao, N. Melman, K. A. Jacobson, J. H. Kime and L. S. Jeong. Design, synthesis and binding affinity of 3'-fluoro analogues of Cl-IB-MECA as adenosine A3 receptor ligands. *Bioorg. Med. Chem. Lett.,* (2003), **13***,* 817–820.
42. I. A. Mikhailopulo, N. E. Poopeiko, T. I. Pricota, G. G. Sivets, E. I. Kvasyuk and E. Clercq. Synthesis and antiviral and cytostatic properties of 3'-deoxy-3'-fluoro- and 2'-azido-3'-fluoro-2',3'-dideoxy-D-ribofuranosides of natural heterocyclic bases. *J. Med. Chem.,* (1991), **34**, 2195-2202.
43. M. F. Ahmed and T. L. Lowary. Glycosylations with 2,3-aziridinofuranose derivatives. *Tetrahedron,* (2013), **69**, 4276-4284.
44. S. De, S. De Jonghe and P. Herdewijn. Synthesis of a 3'-Fluoro-3'-deoxytetrose adenine phosphonate. *J. Org. Chem.,* (2017), **82**, 9464-9478.
45. G. Medgyes and J. Kuszmann. Synthesis of 3-amino-2,3,6-trideoxy-L-lyxo-hexose (daunosamine) hydrochloride from D-glucose. *Carbohydr. Res.,* (1981), **92**, 225-237.
46. J. Shimabukuro, H. Makyio, T. Suzuki, Y. Nishikawa, M. Kawasaki, A. Imamura, H. Ishida, H. Ando, R. Kato and M. Kiso. Synthesis of seleno-fucose compounds and their application to the X-ray structural determination of carbohydrate-lectin complexes using single/multi-wavelength anomalous dispersion phasing. *Bioorg. Med. Chem.,* (2017), **25,** 1132–1142.
47. K. J. Hale, L. Hough, S. Manaviazar and A. Calabrese. An update of the rules for pyranoside sulfonate displacement. *Org. Lett.* (2014), **16**, 4838–4841.
48. M. Giordano and A. Iadonisi. Tin-mediated regioselective benzylation and allylation of polyols: applicability of a catalytic approach under solvent-free conditions. *J. Org. Chem.,* (2014), **79**, 213–222.
49. J. G. Gardiner and E. Percival. 283. Methyl ethers of L-fucose. *J. Chem. Soc.,* (1958), 1414-1418.
50. D. Lamba, A. L. Segre, G. Fabrizi and B. Matsuhiro. Structural analysis of methyl α-L-fucopyranoside by X-ray crystallography, NMR spectroscopy, and molecular mechanics calculations. *Carbohydr. Res.,* (1993), **243**, 217-224.
51. P. J. Card. Fluorinated carbohydrates. Use of (diethylamino)sulfur trifluoride in the synthesis of fluorinated sugars. *J. Org. Chem.*, (1983), **48**, 393-395.

52. P. J. Card and G. S. Reddy. Fluorinated carbohydrates. 2. Selective fluorination of gluco- and mannopyranosides. Use of 2-D NMR for structural assignments. *J. Org. Chem.,* (1983), **48**, 4734-4743.
53. B. K. Gadakh, P. R. Patil, S. Malik and K. P. R. Kartha. Novel selectivity in carbohydrate reactions, IV: DABCO-mediated regioselective primary hydroxyl protection of carbohydrates. *Synth. Commun.,* (2009), **39**, 2430-2438.
54. S. B. Baker. Carbohydrate thioacetals. I. Lead tetraacetate oxidation of D-arabinose derivatives. *J. Am. Chem. Soc.,* (1952), **74**, 827-828.
55. F. W. Lichtenthaler, F. D. Klinger and P. Jarglis. Simple synthesis of (S)-parasorbic acid and other (5S)-hydroxy six-carbon synthons from L-rhamnose. *Carbohydr. Res.,* (1984), **132**, Cl-C4.
56. K. R. Prasad and S. M. Kumar. Total synthesis of (+)-anamarine. *Tetrahedron,* (2014), **70**, 4552-4556.
57. R. E. Ireland and C. S. Wilcox. New and efficient synthesis of 6-deoxy-L-gulose. *J. Org. Chem.,* (1980), **45**, 197-202.
58. T. Řezanka, L. Siristova, K. Melzoch and K. Sigler. N-acylated bacteriohopanehexol-mannosamides from the thermophilic bacterium "Alicyclobacillus acidoterrestris". *Lipids.,* (2011), **46**, 249-261.
59. T. Okada, T. Asawa, Y. Sugiyama, M. Kirihara, T. Iwai and Y. Kimura. Sodium hypochlorite pentahydrate ($NaOCl\cdot 5H_2O$) crystals as an extraordinary oxidant for primary and secondary alcohols. *Synlett*., (2014), **25**, 596-598.
60. M. Attoui and J.-M. Vatèle. TEMPO/NBu_4Br-catalyzed selective alcohol oxidation with periodic acid. *Synlett*., (2014), **25**, 2923-2927.
61. J.-M. Vatèle. $Yb(OTf)_3$-Catalyzed oxidation of alcohols with iodosylbenzene mediated by TEMPO. *Synlett.,* (2006), **13***,* 2055-2058.
62. O. Kölln and H. Redlich. Trimethylene dithioacetals of carbohydrates; part 1: D-ribose, D-allose and D-altrose propane-1,3-diyl dithioacetal and their isopropylidene derivatives. *Synthesis,* (1995), **1995,** 1376-1382.
63. J. S. Elder, J. Mann and E. B. Walsh. Synthesis of a novel prostaglandin H_2 (PGH_2) analogue. *Tetrahedron,* (1985), **41**, 3117-3125.
64. F. Compernolle, G. Joly, K. Peeters, S. Toppet and G. Hoornaert. The synthesis of ester and ketone analogues of 1-deoxynojirimycin and castanospermine. *Tetrahedron,* (1997), **53**, 12739-12754.
65. U. M. Lindström and P. Somfai. Asymmetric synthesis of (+)-1-deoxynojirimycin. *Tetrahedron Lett.,* (1998), **39,** 7173-7176.
66. J. Lehmann and S. Petry. Photolysis of 2-azi-2-deoxy-D-m&no-hexitol and analogous hexitols; the ineffectiveness of certain carbohydrate diazirines as photoaffinity labels. *Carbohydr. Res.,* (1993), **239**, 133-142.
67. A. P. Kozikowski, A. H. Fauq, G. Powis and D. C. Melder. Efficient synthetic routes to fluorinated isosteres of inositol and their effects on cellular growth. *J. Am. Chem. Soc.*, (1990), **112**, 4528-4531.

68. J. Kuszmann and P. Sohár. Synthesis of 1,2:5,6-dianhydro-3,4-dideoxy-erythro and –D-threo-hexitol and their E-3-ene derivatives. *Carbohydr. Res.,* (1980), **83**, 63-72.
69. R. Siedlecka, J. Skarzewski and J. Młochowski. Selective oxidation of primary hydroxy groups in primary-secondary diols. *Tetrahedron Lett.,* (1990), **31**, 2177-2180.
70. M. Adinolfi, G. Barone, A. Iadonisi, L. Mangoni and R. Manna. Synthesis of Caryose, the carbocyclic monosaccharide component of the lipopolysaccharide from *Pseudomonas caryophylli*. *Tetrahedron,* (1997), **53**, 11767-11780.
71. K. Sato, S. Akai, H. Youda, M. Kojima, M. Sakuma, S. Inaba and K. Kurosawa. Practical synthesis of d-[1-13C]Mannose, l-[1-13C] and l-[6-13C]Fucose. *Tetrahedron Lett.,* (2005), **46**, 237-243.
72. K. Matsumoto, K. Koyachi and M. Shindo. Asymmetric total syntheses of xanthatin and 11,13-dihydroxanthatin using a stereocontrolled conjugate allylation to γ-butenolide. *Tetrahedron,* (2013), **69**, 1043-1049.
73. R. Polt, D. Sames and J. Chruma. Glycosidase Inhibitors: Synthesis of enantiomerically pure aza-sugars from schiff base amino esters via tandem reduction-alkenylation and osmylation. *J. Org. Chem.,* (1999), **64**, 6147-6158.
74. W. Yang, S. S. Patil, C.-H. Tsai, C.-H. Lin and J.-M. Fang. The synthesis of L-gulose and L-xylose from D-gluconolactone. *Tetrahedron*, (2002), **58**, 253-259.
75. J. Kuszmann. 1,4-diamino-1,4-dideoxy-D-galactitol and 1,5-diamino-1,5-dideoxy-L-altritol. *Carbohydr. Res.,* (1986), **156**, 25-37.
76. A. A. Bell, R. J. Nash and G. W. J. Fleet. Acetonides of octonolactones. *Tetrahedron: Asymmetry,* (1996), **7**, 595-606.
77. Z. Liu, A. Yoshihara, S. F. Jenkinson, M. R. Wormald, R. J. Estévez, G. W. J. Fleet and K. Izumori. Triacetonide of glucoheptonic acid in the scalable syntheses of D-gulose, 6-deoxy-D-gulose, L-glucose, 6-deoxy-L-glucose and related sugars. *Org. Lett.,* (2016), **18**, 4112-4115.
78. R. Karnekanti, M. Hanumaiah and G. V. M. Sharma. Stereoselective total synthesis of (+)-anamarine and 8-epi-(–)-anamarine from D-mannitol. *Synthesis,* (2015), **47**, 2997–3008.
79. J. B. Rannes, A. Ioannou, S. C. Willies, G. Grogan, C. Behrens, S. L. Flitsch and N. J. Turner. Glycoprotein labeling using engineered variants of Galactose Oxidase obtained by directed evolution. *J. Am. Chem. Soc.,* (2011)**, 133**, 8436-8439.
80. F. Escalettes and N. J. Turner. Directed evolution of Galactose Oxidase: generation of enantioselective secondary alcohol oxidases. *ChemBioChem,* (2008), **9**, 857-860.
81. A. T. Pedersen, W. R. Birmingham, G. Rehn, S. J. Charnock, N. J. Turner and J. M. Woodley. Process requirements of Galactose Oxidase catalyzed oxidation of alcohols. *Org. Process Res. Dev.,* (2015), **19**, 1580-1589.
82. L. H. Sun, I. P. Petrounia, M. Yagasaki, G. Bandara and F. H. Arnold. Expression and stabilization of Galactose Oxidase in *Escherichia coli* by directed evolution. *Protein Eng.,* (2001), **14**, 699-704.

83. S. J. Angyal. The composition of reducing sugars in solution. *Adv. Carbohydr. Chem. Biochem.,* (1984), **42**, 15-68.

CHAPTER 7: CONCLUSIONS

7.1. Conclusions

The following conclusions can be drawn from the work presented in this Thesis:

- The interaction between DC-SIGN and the blood groups A and B is essentially **driven by the Fuc moiety**. By HSQC-titrations, this recognition event has been demonstrated to take place in a specific fashion at the primary binding site (Calcium site 2).
- Both oligosaccharides preserve their respective conformations in the free state upon lectin binding, as assessed by tr-NOESY experiments. These observations agree with the known rigidity already reported for this type of structures. Only minor variations in the glycosidic torsion angles could be detected by MD for the terminal Gal/GalNAc, although they are simply derived from a suitable ring accommodation to the protein surroundings.
- The stability of the CRD is modified as a result of calcium targeting by an entering ligand. Specifically, the **long loop region** that takes part in the structural integrity of the binding site **gains stability**, as proven by HSQC and, indirectly, by DOSY experiments. In addition, the detected conformational changes at the protein backbone are exclusively triggered by the primary sugar interaction. In fact, the set of affected amino acids is similar for every specific ligand, but the magnitude of the chemical shift perturbations varies only depending on whether the primary sugar is Man or Fuc.
- The **non-reducing end (Gal/GalNAc) establishes close contacts with the protein** during the interaction, as evidenced by STD NMR experiments. These contacts favorably contribute to the affinity. According to the performed MD simulations, they basically consist of hydrogen bonds involving several hydroxyl groups (OH3, OH4 and OH6) and more importantly, the hydrophobic shield jointly formed by both non-reducing moieties (Fuc and Gal/GalNAc) around the aliphatic residue Val351. These findings are in accordance with the crucial role that Val351 plays in the recognition of Fuc-containing glycans and could be confirmed by the intermolecular contacts detected in the tr-NOESY.
- Although the binding epitope is the same for both blood antigens, they clearly differ in their dynamic behavior. In particular, the **blood group B exhibits more complex binding dynamics due to the**

existence of an additional binding mode, as evidenced by HSQC-based titrations.

- Gal is also a specific substrate for DC-SIGN, although its affinity is rather low. By using ^{19}F-NMR, Gal has been shown to exclusively interact at the primary binding site through hydroxyl groups OH3 and OH4. Surprisingly, the same conclusion has been reached for Fuc, which cannot be recognized through the OH2-OH3 pair, as reported for langerin or MBL. In this scenario, the **secondary binding mode of the B** antigen that better fulfills the aforementioned observations is that in which the saccharide is **alternatively anchored to the calcium site through OH3 and OH4 of the terminal Gal**. Studies carried out in parallel with the Galili antigen and MD simulations corroborate these findings.
- Regarding the Lewis-type antigens also studied herein, the **LDNF trisaccharide is placed at the binding site in the same way as already described for the Le^x antigen**. In fact, the same secondary contacts with the protein surface are also found for our antigen, namely the accommodation of the GalNAc moiety at the adjacent shallow site and the aliphatic contact between the Fuc moiety and Val351, which in this case is also assisted by the N-acetyl group of the core GlcNAc, as evidenced by STD-NMR and tr-NOESY. The MD simulations actually highlight the importance of such a hydrophobic stabilization mediated by the central residue.
- In contrast, the LDN-DF displays a negligible binding, rather difficult to detect either by HSQC-based titrations and STD-NMR. Consequently, the binding mode has not been entirely delineated, although the MD simulations strongly suggest that **the recognition is exclusively driven by the external Fuc**, while the surrounding sugars (the LDNF portion) is placed away from the protein surface. This is supported by the observed low affinities, above 10 mM.
- The first synthesis of 3-deoxy-3-fluoro-L-Fuc has been achieved by using **a D- to L-sugar conversion strategy**, in 7 steps starting from a cheap commercial D-Glc material. Thestrategy was based in an alcohol inversion step prior to the deoxyfluorination operation, in contrast to the classical methodology that uses straight S_N2 fluorination in a conveniently modified precursor.

7.2. Scientific publications

- **Journal papers:**

L. Unione, A. Gimeno, P. Valverde, I. Calloni, H. Coelho, S. Mirabella, A. Poveda, A. Ardá and J. Jiménez-Barbero. Glycans in infectious diseases. A molecular recognition perspective. *Curr. Med. Chem.,* (2017), **24**, 4057-4080.

P. Valverde, S. Delgado, J. D. Martínez, J.-B. Vendeville, J. Malassis, B. Linclau, N.-C. Reichardt, F. J. Cañada, J. Jiménez-Barbero and A. Ardá. Molecular insights into DC-SIGN binding to self-antigens: the interaction with the blood group A/B antigens. *ACS Chem. Biol.,* (2019), **14**, 1660−1671.

P. Valverde, A. Ardá, N.-C. Reichardt, J. Jiménez- Barbero and A. Gimeno. Glycans in drug discovery. *Medchemcomm.,* (2019), **10**, 1678-1691.

J. D. Martínez, P. Valverde, S. Delgado, C. Romanò, B. Linclau, N.-C. Reichardt, S. Oscarson, A. Ardá, J. Jiménez-Barbero and F. J. Cañada. Unraveling sugar binding modes to DC-SIGN by employing fluorinated carbohydrates. *Molecules,* (2019), **24**, E2337.

P. Valverde, J. I. Quintana, J. I. Santos, A. Ardá, and J. Jiménez-Barbero. Novel NMR avenues to explore the conformation and interactions of glycans. *ACS Omega.,* (2019), **4**, 13618-13630.

A. Gimeno, S. Delgado, P. Valverde, S. Bertuzzi, M. A. Berbís, J. Echavarren, A. Lacetera, S. Martín-Santamaría, A. Surolia, F. J. Cañada, J. Jiménez-Barbero and A. Ardá. Minimizing the entropy penalty for ligand binding: Lessons from the molecular recognition of the histo blood-group antigens by human galectin-3. *Angew. Chem. Int. Ed.,* (2019), **58**, 7268 –7272.

A. Gimeno, P. Valverde, A. Ardá and Jesús Jiménez-Barbero. Glycan structures and their interactions with proteins. A NMR view. *Curr. Opin. Struct. Biol.,* (2020), **62**, 22–30.

- **Book chapters:**

A. Ardá, H. Coelho, B. F. de Toro, S. Galante, A. Gimeno, A. Poveda, J. Sastre, L. Unione, P. Valverde, F. J. Cañada and J. Jiménez-Barbero. Recent advances in the application of NMR methods to uncover the conformation and recognition features of glycans. In *Carbohydr. Chem.,* (2017), **42**, 47-82.

- **Other publications in preparation:**

A. D. Srivastava, L. Unione, P. Valverde, M. A. Wolfert, A. Ardá, J. Jiménez-Barbero and G.-J. Boons. Synthesis and Molecular Recognition of fucosylated glycans of the parasitic worm *S. mansoni* by the innate immune receptor DC-SIGN. *Submitted.*

P. Valverde, J.-B. Vendeville, A. Mattey, K. Huonnic, M. Light, J. Jimenez-Barbero, N. Turner, S. Flitsch and B. Linclau. A chemoenzymatic synthesis of 3-Deoxy-3-fluoro-L-fucose from methyl-α-D-glucoside. *In preparation.*

7.3. Contributions to congresses

Poster: Structural characterization of the TubR CST190 protein by Nuclear Magnetic Resonance. XII Jornadas de Carbohidratos. March 2016, Madrid, Spain.

Poster: NMR studies of the specific Interaction of DC-SIGN with different oligosaccharides. 8th GERMN / 5th Iberian NMR Meeting. June 2016, Valencia, Spain.

Poster: NMR studies of the specific Interaction of DC-SIGN with different oligosaccharides. 10th "NMR as a Tool for Biology" Conference. January 2017, Paris, France.

Poster: NMR studies of the specific Interaction of DC-SIGN with different oligosaccharides. XXXVI Reunión Bienal de la Real Sociedad Española de Química. June 2017, Sitges, Spain.

Poster: NMR studies of the specific Interaction of DC-SIGN with fucosylated ligands. 29TH International Carbohydrate Symposium. July 2018, Lisbon, Portugal.

Oral communication: Recognition of Carbohydrate Antigens. Glycovax Training Event IV. March 2018, Bilbao, Spain.

Poster: NMR studies of the specific Interaction of DC-SIGN with fucosylated ligands. X Spanish Drug Discovery Network Meeting. November 2018, Bilbao, Spain.

Flash communication: Molecular insights into DC-SIGN binding to self antigens: the interaction with the blood group A/B antigens. XXXVII Reunión Bienal de la Real Sociedad Española de Química. May 2019, Donostia, Spain.

Flash communication: Molecular insights into DC-SIGN binding to self antigens: the interaction with the blood group A/B antigens. Introductory Workshop on Biomedical Glycoscience. June 2019, Donostia, Spain.

Oral communication: Molecular insights into DC-SIGN binding to self antigens: the interaction with the blood group A/B antigens. XX European Carbohydrate Symposium (Eurocarb XX). July 2019, Leiden, Netherlands.

Flash communication: Molecular insights into DC-SIGN binding to self antigens: the interaction with the blood group A/B antigens. February 2020, Granada, Spain.